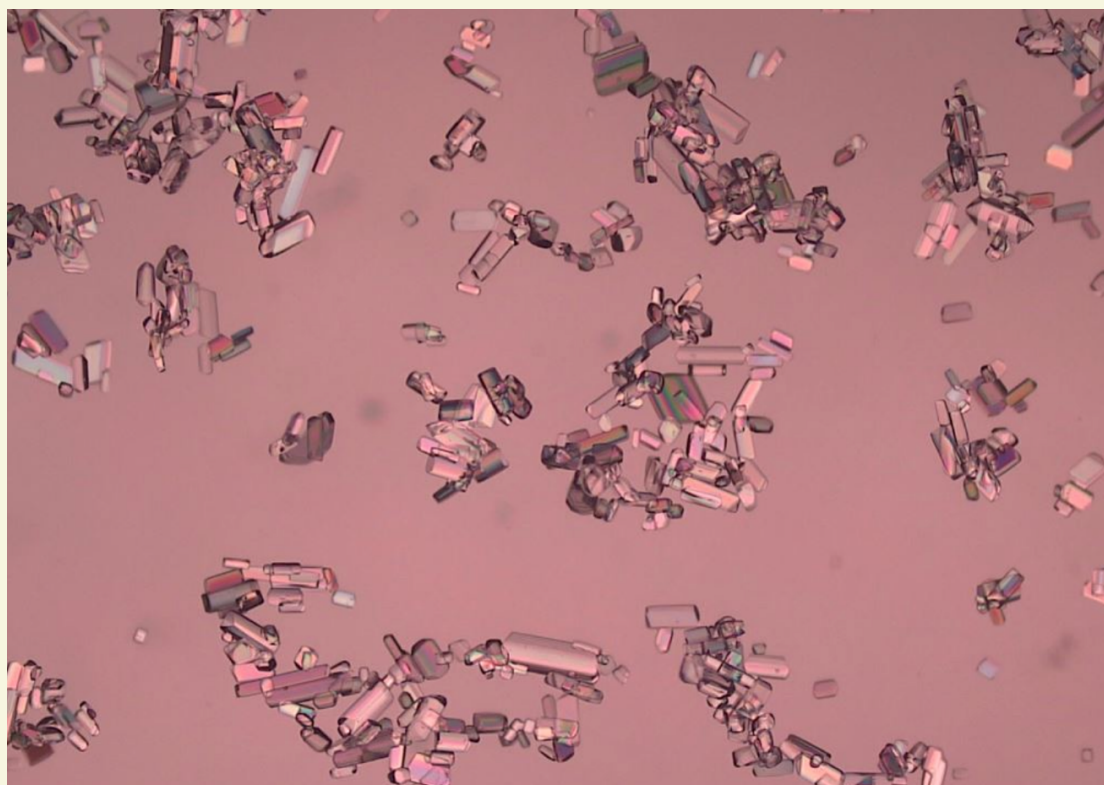


CHEMICAL ENGINEERING RESEARCH

*Reports of the 4th year research projects
in the Department of Chemical
Engineering at Imperial College London*

VOLUME 8



Edited by Erich A. Müller

February, 2026

CHEMICAL ENGINEERING RESEARCH

Reports of the 4th year research projects
in the Department of Chemical Engineering
at Imperial College London

Edited by Erich A. Müller

Volume 8

2026



© The Author(s) 2026.

Published by Imperial College London, London SW7 2AZ, UK

Contact details – e.muller@imperial.ac.uk

This book is a compilation of manuscripts curated from a teaching assignment in the 4th year's Chemical Engineering CENG70001 course (Advanced Chemical Engineering Practice Research Project). Copyright in each paper rests with its authors. The author of each paper appearing in this book is solely responsible for the content thereof; the inclusion of a paper in the book shall not constitute or be deemed to constitute any representation by Imperial College London that the data presented therein are correct or sufficient to support the conclusions reached or that the experiment design or methodology is adequate.



The book is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0). Under this licence, you may copy and redistribute the material in any medium or format on the condition that: you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law. <https://creativecommons.org/licenses/by-nc-nd/3.0/>.

The information contained in this publication is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall Imperial College London be liable for damages arising from its use.

E-ISBN 978-1-9160050-7-5

First published in 2026

Preface

This volume of *Chemical Engineering Research* collects the unedited research project reports written by 4th-year undergraduates (Class of 2026) of the M.Eng. course on Chemical Engineering in the Department of Chemical Engineering at Imperial College London. The research project spans one term (Autumn) during the last year of the career. It emphasises independence, the ability to plan and pursue original project work for an extended period, produce a high-quality report, and present the work to an audience using appropriate visual aids. Students are also expected to produce a literature review and situate their work in the context of prior art. The papers presented showcase the diversity and depth of some of the research streams in the department, but only touch on a small number of research groups and interests. For a complete description of the research at the department, the reader is referred to the departmental website¹.

The papers are presented in no particular order, each identified by a manuscript number. Some papers refer to appendices and/or supplementary materials that are too lengthy to include here. These files can be obtained directly from the supervisors (see the supervisor index at the end of the book). Some reports are missing and are under embargo as they contain confidential information.

The cover figure shows a microscopic (10X) image of alanine crystals in deionized water (from the work of Adam Adan and Aliff Rafiq, manuscript 13).

London, February 2026

¹ <https://www.imperial.ac.uk/chemical-engineering>

Title Index

paper	Title	page
1	Dynamic Viscosity Prediction of Aromatics and Naphthenes using Group Contribution SAFT	1
2	Optimisation and Evaluation of a Lipstick-Based Colorimetric Biosensor for Salivary pH Analysis Using Red Cabbage Powder	11
3	Automated Process Synthesis: Implementation and Evaluation of a Genetic Algorithm for Flowsheet Generation and Optimisation	21
4	Evaluation of the Natural Ortho-Para Hydrogen Conversion Kinetics	31
5	Hybrid Normalised Advantage Function Reinforcement Learning for Energy-Efficient Switching Control in Temperature Swing Adsorption Systems	41
6	Inverse Engineering of Swift-Hohenberg Patterns via Physics-Informed Machine Learning	51
7	Predictive Viscosity Model for Sustainable Aviation Fuel	61
8	Optimising Hydrogen Deployment for Long-Duration Energy Storage in Future UK Electricity Networks	71
9	Optimizing Copper (Cu) Electrodeposition via Different Control Strategies	81
10	Process Design of a Hydrogen Liquefaction Plant Operated with Intermittent Wind Electrolysis	91
11	Investigation of different factors affecting CO ₂ capture by CALF-20	101
12	Screening the chemical space: An ML approach to predicting the cost of chemical compounds	(*)
13	Investigating The Effects of Protecting Groups on the Thermodynamics and Kinetics of Alanine	110

paper	Title	page
14	Evaluating Actor-Critic Reinforcement Learning Algorithms for Process Control Applications in a CSTR	120
15	Extracting Chitin from Mushrooms with Ionic Liquids	130
16	Life Cycle Assessment and Technical Economic Analysis of Polyethylene Terephthalate Recycling Processes	(*)
17	Cost Optimisation of Hydrogen Integration in the ASEAN Power Grid Using TZ-APG PyPSA Model	140
18	Dynamic Fouling Modelling in Skimmed Milk Powder Manufacturing	150
19	Graphene-on-Foil Sheets for Electrochemical and Photoelectrochemical Water Splitting	(*)
20	Assembling Motile Biohybrids via Coordinate Bonding Between DGS-NTA(Ni) Lipids and 6-Histidine	162
21	Fabrication of Composite SU-102@Anodic Aluminium Oxide Membranes with Characterisation and Anionic Dye Rejection Properties	172
22	Techno-Economic Analysis of Crystallisation for the Continuous Production of Paracetamol with Solvent Recycling	183
23	Benchmarking Automated, High-throughput Adams' Fusion Synthesis of IrO ₂ OER Electrocatalysts	193
24	Evaluation of PEGDA-based microgels as a platform for tattoo biosensing	203
25	Multi-solvent computer-aided molecular and process design framework to predict VLE properties	213
26	CFD of Blood Flow in Marfan Patients: A Haemodynamic Comparison between Medical Management and Aortic Root Replacement Surgery	223
27	Numerical Modelling and Process Evaluation of Structural Transitional Adsorbents for Adsorption-based Carbon Capture	233
28	k-BAAM: A kinetics-augmented batch adsorber analogue model for sorbent screening in pressure-vacuum swing adsorption process for point-source CO ₂ capture	243

paper	Title	page
29	Reaction Modelling and Geometry Optimisation for Multi-tubular Membrane Catalytic Ammonia Cracker	252
30	Computational modelling of thrombolysis: Effect of varying therapeutic drug and characteristics of the blood clot	261
31	Bayesian Optimization in Latent Variable Models for Expensive Black-Box Problems	271
32	PEG-based pH/redox Dual-responsive Nanogel for Oral Drug Delivery	(*)
33	Application and Optimisation of Symbolic Regression in Crystallisation Modelling	281
34	Bipolar membrane Electrodialysis (BMED) for PET recycling	(*)
35	Lignin-Directed Pore-Mouth Engineering Enables Tunable Sodium Storage in Hard Carbon Anodes	290
36	Targeted nanomedicine for effective thrombolysis	299
37	Modelling of 3D Solute Transport in Disordered Porous Media	309
38	Mathematical Modeling of Spray-Based Microalgae Culture Systems in Column Photobioreactors	319
39	Fabrication and Characterisation of Matrimid 5218-CPOC-301 Mixed-Matrix Membranes	329
40	Optimisation of the Pd/NF Catalyst for Glycerol Electrochemical Oxidation Reaction	338
41	Characterising SAF: Evaluating the Physicochemical Properties of Sulphonated Alkyl Furoates	348
42	The Use of Neural Networks in the Prediction of Thermal Conductivity of Organic Liquids: Key Features and the Added Value of Molecular Fingerprints	358
43	Optimising electrochemical biomass waste upcycling with automated experiments	367

paper	Title	page
44	Enhanced Fresh Water Desalination Using Barium Bismuth Oxide with Efficient Energy Management	377
45	Temperature Dependence of Early-Age CO ₂ Sequestration in Coke-Modified Concrete	387
46	Superhydrophobic Modification of Different Cotton Types for Oil-Water Separation	397
47	Dilatational Surface Viscous Effects on Bubble Bursting at Low Bond Number	406
48	Machine Learning Enhanced Geometric Optimisation of Crystallisation Reactors	417
49	Carbon Capture on Waste-to-Energy Plant	427
50	Assessing Flexibility in the Haber-Bosch Process Under Variable Throughput	437
51	Hydrothermal Carbonisation for Sodium-Ion Battery Materials: Towards Continuous Flow Production	(*)
52	Investigating the Impact of Electrode Material on the Synthesis of Carbon Nanotubes via Molten Salt Electrolysis	447
53	A Systematic Techno-economic Review of Carbon Capture and Storage Technologies in the Global Cement Industry	457
54	Evaluating Strategies to Reduce Experimental Load for Design Space Identification	469
55	Modelling of Integrated Photovoltaic-Photoelectrochemical System for Hydrogen Production	479
56	Virus-mimicking lipid nanoparticles for intracellular drug delivery	(*)
57	Biodegradable Microcapsule Formation for Analgesic Delivery via Acoustic Levitation	489
58	Defining Structural Nativeness For De Novo Antibody Design	499

paper	Title	page
59	Atmospheric Plasma-Induced Wrinkling as a Scalable Surface Patterning Technique	509
60	Noble-Metal-Free CdS/MoP Photocatalyst for Enhanced Selective Conversion of Methanol to Ethylene Glycol under Visible Light	519
61	An Integrated Dynamic Vapour Sorption and Low-Field NMR Platform for Resolving Microscopic Water Environments	(*)
62	Repeatability of ionic liquid treatment for cotton and dye recycling process	527
63	Exploring the effect of amphiphile morphology of APGs on interfacial tension	537
64	The Effects of Amino Acid Impurities and Well Conditions on Vancomycin Crystallisation	547
65	Orthogonal Collocation on Finite Elements Method for Simulation of Membrane Contactors in CO ₂ Capture	555
66	Extending Liquid Phase Oligonucleotide Synthesis to a Two-Stage Membrane Cascade: Dynamic Modelling and Feasible Operating Region Analysis	565
67	Techno-economic Analysis of Hybrid Energy Systems for Powering Hyper-scale Data-centres in the UK	575
68	Development and Characterisation of Experimental Standards for Biological Measurements	(*)
69	Global Sensitivity Analysis of Economic Parameters in CAR T-Cell Therapy Manufacturing	585

(*) These papers have been removed at the request of the authors and/or supervisors

An Author and Supervisor index is at the end of the book.

Dynamic Viscosity Prediction of Aromatics and Naphthenes using Group Contribution SAFT

Alfred Hoyle and Joe Wood

Department of Chemical Engineering, Imperial College London, U.K.

Abstract: In this work, a new group-contribution entropy-scaling method is proposed for predicting the dynamic viscosity of n-alkyl-benzenes and naphthenes. The approach modifies the entropy-scaling framework of Bell and Yang [Yang, X. et al., *J. Chem. Eng. Data*, 2021, 66, 1385–1398] by replacing molecule-specific parameters with group-dependent ones. Entropy is obtained using the SAFT- γ Mie group-contribution equation of state, and group-specific parameters are estimated by optimising the agreement between experimental measurements and the aggregate of the corresponding group contributions. The parameter for any given molecule is then computed as the weighted sum of all group parameters, scaled by each group's occurrence. This formulation enables viscosity prediction without requiring pure-component parameters and therefore increases the method's applicability across diverse molecular structures. Temperature-dependent and chain-length-dependent correction terms were also developed, allowing the model to maintain accuracy across homologous series and a broad range of thermodynamic conditions. Applied to a dataset of alkyl-cyclic and alkyl-aromatic compounds with C1–C12 alkyl chains, predictions using this method achieves mean AAD% values of 6.40 for naphthenes and 9.11 for aromatics. These results demonstrate strong predictive capability while significantly reducing parameter demands. Overall, the framework offers a scalable and transferable approach to viscosity prediction, with clear potential for extension to mixtures, elevated pressures, and more complex hydrocarbon families

1. Introduction

Climate change is one of the most serious global issues. To combat this, many countries, such as the UK, have set emissions targets for net zero by 2050¹. The aviation industry accounts for 2.5% of global CO₂ emissions and is therefore a significant focus for emissions reduction, particularly through the introduction of sustainable aviation fuels (SAF). These fuels are produced from a variety of sources, most commonly biowaste, natural oils, and sugars². Although the composition of these fuels varies considerably, the four main groups of compounds found in SAF are n-alkanes, branched alkanes, naphthenes and aromatics. Figure 1.1 shows the composition of one such blend³.

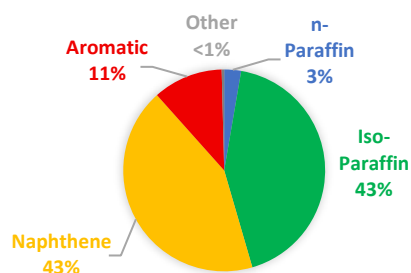


Figure 1.1 Composition of FT-SPK, a common SAF blend

Aviation fuels must have suitable properties to be viable for use, including viscosity. Experimental approaches to SAF development are expensive and time-consuming: to improve efficiency, this work aims to produce a group-contribution model for predicting n-alkane, aromatic and naphthene viscosities, selected for their prominence in existing SAF, like FT-SPK (figure 1.1).

2. Background

2.1 Cubic Equations of State

Cubic equations of state are relied on heavily in modern industry to obtain thermodynamic properties, due to their simplicity. These are based on the van der Waals equation (2.1), which models molecules as hard spheres.

$$p = \frac{RT}{v-b} - \frac{a}{v^2} \quad (2.1)$$

Where p is pressure (Pa), T is temperature (K), R is the molar gas constant ($\text{J K}^{-1} \text{mol}^{-1}$) and v is molar volume ($\text{m}^3 \text{mol}^{-1}$). a ($\text{Pa m}^6 \text{mol}^{-2}$) and b ($\text{m}^3 \text{mol}^{-1}$), in this instance, represent the attractive forces between molecules and the finite volume of the hard spheres, respectively. This equation was further developed into the Soave-Redlich-Kwong equation (2.2) and Peng Robinson equation (2.3). These introduce a third term, $\alpha(T)$, to account for temperature dependence of the attractive forces.

$$p = \frac{RT}{v(v-b)} - \frac{a\alpha(T)}{v(v+b)} \quad (2.2)$$

$$p = \frac{RT}{v(v-b)} - \frac{a\alpha(T)}{(v^2 + 2bv - b^2)} \quad (2.3)$$

While these prove more accurate for most hydrocarbons found in SAF, they are highly empirical and, therefore, non-predictive - parameters a , b and α are fitted per compound.

Secondly, due to the large change in molar volume around the phase change from gas to liquid, these equations often struggle to predict liquid phase properties. Various methods, such as NRTL and UNIFAC, exist for this. However, for a more comprehensive EOS description, Statistical Associating Fluid Theory (SAFT) provides a suitable alternative.

2.2 Statistical Associating Fluid Theory

SAFT is a statistical-thermodynamics-based framework used to determine the Helmholtz free energy, A , of a fluid. This is achieved by summing the contributions to the free energy from the ideal model, the monomer, chain, and association contributions, as seen below:

$$A = A_{\text{ideal}} + A_{\text{mono}} + A_{\text{chain}} + A_{\text{assoc}} \quad (2.4)$$

The residual Helmholtz energy is further extracted from this:

$$A_{\text{res}} = A - A_{\text{ideal}} = A_{\text{mono}} + A_{\text{chain}} + A_{\text{assoc}} \quad (2.5)$$

The ideal term, A_{ideal} , is simply the ideal gas free energy which can be obtained from integrating the molecular partition function, q . The monomer term represents the adjustment for hard spheres like those present in the van der Waals equation and is also often referred to as the HC or hard sphere term. The chain term adjusts for the fact that the molecules are non-spherical; for instance, long chain n-alkanes. Lastly, the association term accounts for interactions between molecules depending on orientation, such as hydrogen bonds in water.

SAFT methods can broadly be divided into homogeneous and heterogeneous types. Homogeneous SAFT methods, such as PC-SAFT, treat molecules explicitly, using parameters fitted to known compounds where data is available. While highly accurate and physics based, it is not applicable for predicting “unseen” molecules. For this reason, this work focuses on heterogeneous methods, such as SAFT- γ Mie. These methods model each molecule as a combination of “groups” such as CH_3 . For instance, isopentane would be modelled as 1 CH , 1 CH_2 and 3 CH_3 groups, shown in figure 2.1. This allows for prediction and scaling of new compounds.

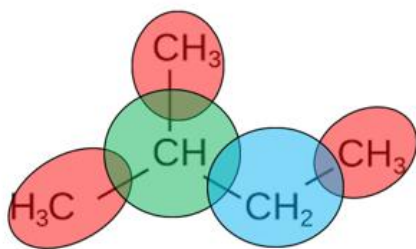


Figure 2.1 Group contribution representation of isopentane

Residual entropy can be simply calculated as the derivative of the free energy with respect to temperature at constant volume and moles:

$$s_{\text{res}} = - \left(\frac{\partial A_{\text{res}}}{\partial T} \right)_{v,n} \quad (2.6)$$

This residual entropy can be used for viscosity calculations using various **entropy scaling methods**.

2.3 Lötgering-Lin and Gross Method

Lötgering-Lin and Gross⁴ developed a multiplicative entropy-scaling method which uses the residual entropy obtained from PCP-SAFT (a heterogeneous SAFT model) to define an entropy based residual viscosity term, η^* . This is then combined with a dilute gas viscosity, obtained from the Chapman-Enskog equation. This is a function of several parameters obtained from PCP-SAFT, such as the number of each group, m_{gc} , and the group segment diameter, σ_{gc} , seen below.

$$\eta_{CE} = \frac{5}{16} \frac{\sqrt{M k_b T / (m_{gc} N_A \pi)}}{\sigma_{gc}^2 \Omega_{gc}^{(2,2)}} \quad (2.7)$$

Where η_{CE} represents the Chapman Enskog viscosity, M refers to the molecular mass (g mol^{-1}), k_b is the Boltzmann constant (J K^{-1}), T is temperature (K), N_A is Avogadro's number and $\Omega_{gc}^{(2,2)}$ is the reduced collision integral (obtained from literature⁵). The residual viscosity, η^* , is then calculated as the exponent of a third degree polynomial of the residual entropy.

$$\ln \eta^* = A + Bz + Cz^2 + Dz^3 \quad (2.8)$$

Where

$$z = \frac{s_{\text{res}}}{k_B m_{gc}} \quad (2.9)$$

The product of these two viscosities gives the total viscosity of the compound

$$\eta = \eta_{CE} \eta^* \quad (2.10)$$

2.3 Bell and Yang Method

An alternative method is outlined by Bell and Yang⁶, in which the residual viscosity term summed with the dilute gas viscosity in a linear fashion, rather than taking the product. This residual viscosity is calculated as a function of the reduced residual viscosity, η_{res}^+ , and the number density ρ_N .

$$\ln(\eta_{\text{res}}^+ + 1) = n_1 \left(\frac{s^+}{\xi} \right)^{1.8} + n_2 \left(\frac{s^+}{\xi} \right)^{2.4} + n_3 \left(\frac{s^+}{\xi} \right)^{2.8} \quad (2.11)$$

Where

$$s^+ = - \frac{s_{\text{res}}}{R} \quad (2.12)$$

In equation (2.11), n_i and ξ are molecule specific constants, or may be given for specific groups of compounds, such as noble gases or aromatics. The residual viscosity is then as follows:

$$\eta_{\text{res}} = \eta_{\text{res}}^+ \frac{\rho_N \sqrt{M k_B T}}{(s^+)^{2/3}} \quad (2.13)$$

The dilute gas viscosity term, $\eta_{\rho \rightarrow 0}$, is equivalent to the Chapman Enskog viscosity term from equation (2.7). Viscosity is then given by:

$$\eta = \eta_{\rho \rightarrow 0} + \eta_{\text{res}} \quad (2.14)$$

3. Methods

3.1 Method Selection

To determine the method to continue with for further modelling, various initial tests were conducted on a variety of n-alkanes. Both methods performed well with medium length (C5-C8) alkanes, but, with larger molecules, the Lötgering-Lin method's multiplicative term does not adequately compensate for the increase in viscosity.

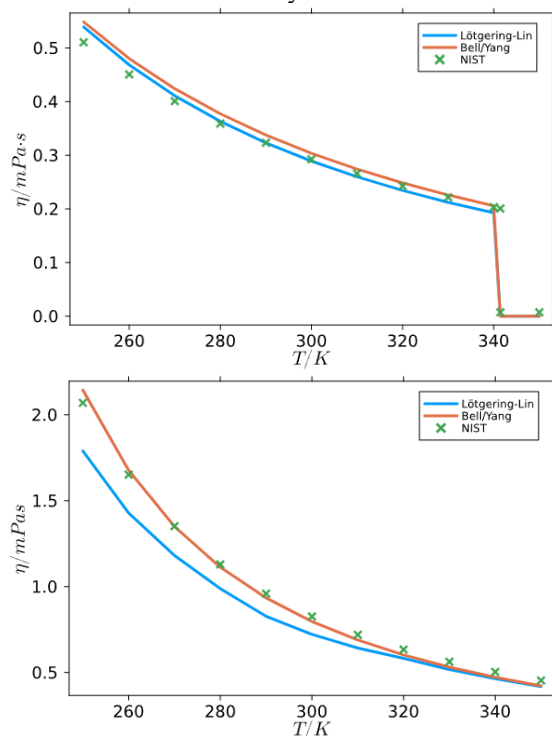


Figure 3.1 Comparison of Bell/Yang and Lötgering-Lin Viscosity Models for n-hexane (above) and n-decane (below)

The Bell and Yang method maintains accuracy, and due to the large size of the naphthenes and aromatic molecules found in SAF, it was the chosen method to base further experiments on. Furthermore, as will be described in section 3.3, the parameters for the Bell and Yang method can be modified to represent group contribution rather than molecule specific behaviours.

3.2 Parameter Selection

Equation 2.11 contains four parameters, outlined by Bell & Yang⁶. In their work, two approaches are taken: firstly, the use of a molecule-specific ξ with molecule-type-specific ' n ' coefficients (i.e. 'group 4' refers to aromatics); secondly, the use of a generalised $\xi=1$ with molecule-specific ' n ' values. In either case, at least one of the two parameter types have been fitted to existing data. To produce a model that can be applied to unseen datasets, allowing predictive capabilities, one must select which of these parameters to change.

A set of optimisations on the viscosity of benzene, obtained from NIST⁷, were carried out in three scenarios, each using four sets of initialisations, to confirm the validity of the optimisation sequence.

Three scenarios:

- 1) Optimising ξ with constant ' n '
- 2) Optimising ' n ' with constant ξ
- 3) Optimising both ξ and ' n ' simultaneously

Table 3.1 Molecule- and Group- Specific Parameters for Benzene from Bell & Yang⁶

	ξ	n_1	n_2	n_3
Benzene	0.838	0.252	-0.043	0.00042
Group 4	-----	0.306	-0.128	0.02866

Table 3.2 Parameter combinations used for each initialisation in the selection process

	$\xi = 1$	$\xi = \text{Benzene}$
$n = \text{Benzene}$	A	C
$n = \text{Group 4}$	B	D

To produce the quantities of importance, i.e. residual entropy and density, the Clapeyron.jl package was utilised in the Julia environment, making use of SAFT- γ Mie to maintain group-contribution applicability of the results. Nelder-Mead optimisations on Least Squares objective functions, in Optim.jl, were used for fitting parameters.

3.3 Modifying the Bell-Yang Equation

Following the methodology described in section 3.2, further work focused on finding appropriate group-contribution functions to approximate a molecule's ξ using a universal set of n coefficients (taken as an average of group coefficients from Bell & Yang's work⁶ to 4 significant figures).

Table 3.3 Global 'n' Coefficients used in modelling

	n_1	n_2	n_3
Global Value	0.3014	-0.1193	0.02531

The first - and most simple – proposed function for ζ is a linear combination of segment ζ values weighted by their occurrence in the molecule:

$$\zeta_{\text{total}} = \sum n_i \zeta_i \quad (3.1)$$

where n_i is the occurrence of a group 'i', and ζ_i is the group contribution ζ for that group.

Initial attempts use a systematic approach to finding segment values of ζ for the groups of interest (aromatics, naphthenes, alkanes). For example, two n-alkanes (A1 and A2) have their total ζ found by Least-Squares optimisation, and the set of linear equations is solved simultaneously to find appropriate values of ζ_{CH_2} and ζ_{CH_3} .

$$n_{\text{A1,CH}_2} \zeta_{\text{CH}_2} + n_{\text{A1,CH}_3} \zeta_{\text{CH}_3} = \zeta_{\text{A1}} \quad (3.2)$$

$$n_{\text{A2,CH}_2} \zeta_{\text{CH}_2} + n_{\text{A2,CH}_3} \zeta_{\text{CH}_3} = \zeta_{\text{A2}} \quad (3.3)$$

Equations 3.2 & 3.3, examples for the systematic determination of Group Contribution (GC) ζ values.

where $n_{X,Y}$ is the number of segments of type 'Y' in molecule 'X', ζ_Y is the GC value corresponding to segment 'Y', and ζ_X is the optimised value of ζ for molecule 'X'.

The same equation was then tested with a batch-wise optimisation of ζ_i for all training-set molecules, producing a distinct set of values. Experimental data for training was obtained from literature⁷⁻⁸.

Table 3.1 Training set molecules

TRAINING SET	Propyl-benzene	Octane
Benzene	Butyl-benzene	Nonane
Toluene	Pentyl-benzene	Decane
P-Xylene	Hexane	Dodecane
Ethyl-benzene	Heptane	

Table 3.2 Testing set of molecules

TESTING SET	Trimethyl-benzene	Hexyl-benzene
Heptyl-benzene	Octyl-benzene	Dodecyl-benzene

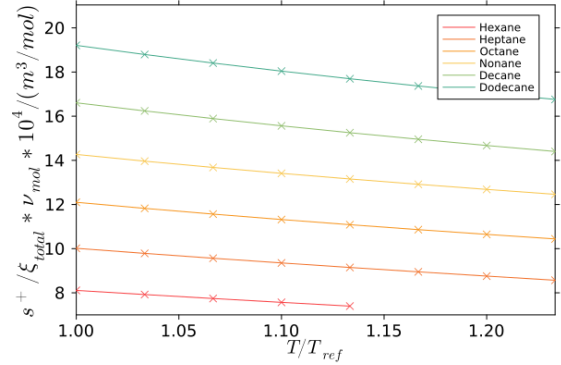
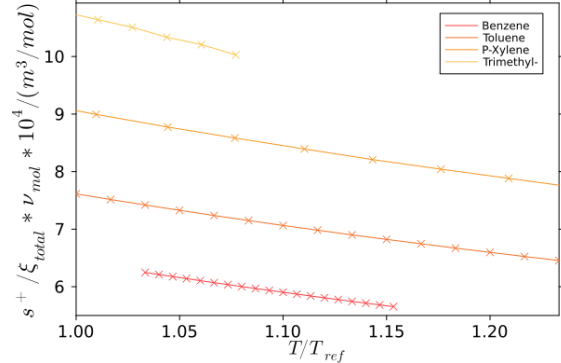
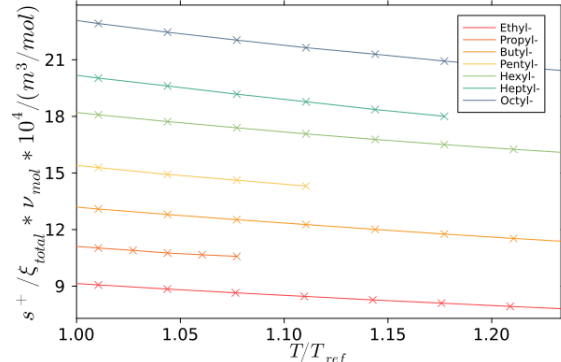
The second proposed function for ζ_{total} introduces a scaling factor based on the total size of the molecule. The existing model, where ζ_{total} increases with increasing molecular size, contradicts known trends in viscosity. With a greater ζ_{total} , the viscosity decreases – this adjustment helps to counteract this effect. For the sake of interpretation and non-dimensionality, the total factor applied was that of an individual segment's effective volume

contribution to the total molecule's effective volume: as all terms in the numerator are constants from SAFT- γ Mie, when optimised, they only serve to change the size of each segment value (ζ_i) without impacting ζ_{total} . The following two equations are thus mathematically equivalent in their output but unique in their definitions of ζ_i . Note that any values of ζ_i for this model are found using Equation (3.4).

$$\zeta_{\text{total}} = \sum n_i \zeta_i \frac{s_i \sigma_i}{\sum n_i s_i \sigma_i^3} \quad (3.4)$$

$$\zeta_{\text{total}} = \sum n_i \zeta_i \frac{1}{\sum n_i s_i \sigma_i^3} \quad (3.5)$$

where n_i is the occurrence of segment 'i', ζ_i is the group-contribution of segment 'i' to the total value, s_i is the SAFT shape-factor of segment 'i', and σ_i is the SAFT diameter of segment 'i'.

**Figure 3.5** Values of $(s^+/\xi)_{\text{Total}} \nu * 10^4$ against Relative Temperature for n-alkanes, $T_{\text{ref}} = 300\text{K}$, $P = 1\text{atm}$ **Figure 3.6** Values of $(s^+/\xi)_{\text{Total}} \nu * 10^4$ against Relative (K) for methyl-benzenes, $T_{\text{ref}} = 300\text{K}$, $P = 1\text{atm}$ **Figure 3.7** Values of $(s^+/\xi)_{\text{Total}} \nu * 10^4$ against Relative Temperature for n-alkyl-benzenes, $T_{\text{ref}} = 300\text{K}$, $P = 1\text{atm}$

The third model uses the above figures to develop a molar density scaling model, showing a mostly linear relationship between the y-axis and relative temperature ($T_{ref} = 300K$), as well as consistent distances between lines (with exceptions, i.e. hexylbenzene). It can be seen that gradients with temperature tend to steepen with larger molecules and, as a result, the following relationship was determined:

$$\frac{-s_{mol}^{res} 10^4}{R \xi_{tot}} = \sum n_i (\xi_{i,n} - \xi_{i,T} \frac{T}{T_{ref}}) \quad (3.6)$$

where s_{mol}^{res} is the residual entropy ($J mol^{-1}$), R is the gas constant ($J K^{-1} mol^{-1}$), v_{mol} is the molar specific volume ($m^3 mol^{-1}$), n_i is the occurrence of segment 'i' in the molecule, $\xi_{i,n}$ is the 'i'th segments temperature-independent contribution, and $\xi_{i,T}$ is the temperature-dependent group-contribution factor.

Equation (3.6) can be rearranged to find ξ_{total} as a function of other variables as follows:

$$\xi_{tot} = \frac{-s_{mol}^{res} 10^4}{R \sum n_i (\xi_{i,n} - \xi_{i,T} \frac{T}{T_{ref}})} \quad (3.7)$$

which subsequently results in the cancellation of the residual entropy term, s^+ , in the Bell-Yang equation for reduced viscosity:

$$\frac{s^+}{\xi} \rightarrow \frac{\sum n_i (\xi_{i,n} - \xi_{i,T} \frac{T}{T_{ref}})}{v_{mol} 10^4} \quad (3.8)$$

and ultimately leads to a (molar) density-scaling based approach on viscosity, including a term for temperature dependence based on the groups present.

3.4 Temperature Dependence

Aside from the last of the aforementioned models, there has been no inclusion of temperature dependence on ξ_{total} . This sub-section describes the main method attempted for incorporating this into the existing models, without changing their form. Again, dependencies on the optimal values of ξ_{total} with different variables were observed and used for the development of the model. In order to keep the temperature-dependent modelling as universally applicable as possible, and to maintain non-dimensionality of ξ (a failing of the density-scaled temperature-dependent model), all values are calculated with respect to a reference point at 300K and 1atm.

$$X_{rel} = \frac{X}{X_{|T=300K, P=1atm}} \quad (3.9)$$

where X is any variable

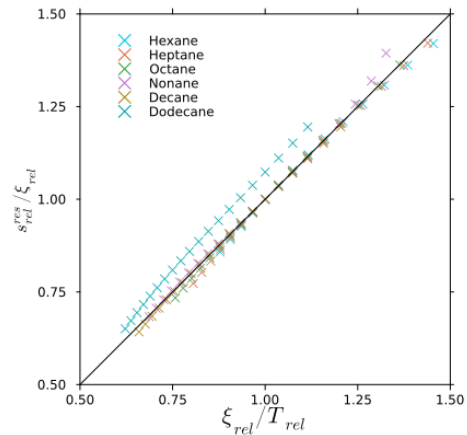


Figure 3.10 Plot of $s_{rel}^{res} / \xi_{rel}$ against ξ_{rel} / T_{rel} for determining a generalised temperature relationship using n-alkanes

For n-alkanes with $C < 10$, a strong linear relationship is observed between the groups on each axis. This results in the following temperature dependent equation for adjusting our existing temperature-independent ξ relationships:

$$\xi_{total}(T) = \xi_{total|T=300K} \sqrt{\frac{s_{res} T}{s_{res|T=300K} T_{ref}}} \quad (3.10)$$

To apply this method appropriately, only data at $T=300K$ is needed. This method has been used for larger molecules, like dodecane, despite their deviation from the $y=x$ line in figure 3.10, to test general applicability.

3.5 Chain Length Dependence for Cyclic and Aromatic Connective Carbons

To further the depth of our model in accounting for the complex shapes of naphthenes and aromatics, adjustments of ξ_i values for aCCH2 and cCH, outlined in figure 3.11, were explored.

In each instance, the contribution of the highlighted groups is affected by the length of the attached chain. Longer chains result in greater occlusion of these joining groups, and, as a result, the model should reflect a decrease in this group's ξ_i .

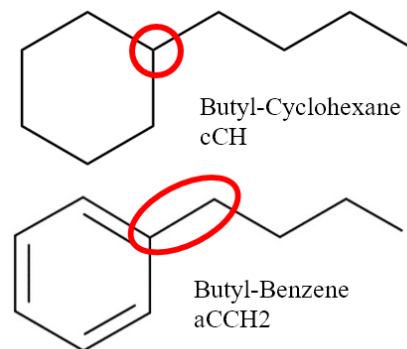


Figure 3.11 Visualisations of the unique groups (cCH & aCCH2) in n-alkyl-naphthenes and n-alkyl-benzenes

Two methods of adjustment were trialled:

- 1) Adjusting the connective carbon's ξ_i as a function of chain length,

$$\xi_{cCH} = \xi_{cCH}(\text{chain length}) \quad (3.10)$$

where n is chain length

- 2) Adjusting the n -alkyl ξ_i as a function of proximity to the connective carbon,

$$\xi_{\text{chain}} = \sum_{n=1}^{\text{chain length}} \alpha_n \xi_{\text{aCCH}_2} + (1 - \alpha_n) \xi_{\text{alkyl}} \quad (3.11)$$

where α_n is a function of the group's distance from the aCCH₂ group such that $\alpha_n \in [0,1]$, and 'n' is, in this case, each alkyl (CH₂, CH₃) group's distance from aCCH₂, in numbers of C.

For the second of these, a number of different decreasing functions for α_n were used and compared.

$$\alpha_n^1 = \frac{1}{n} \quad (3.12)$$

$$\alpha_n^2 = \frac{1}{n^2} \quad (3.13)$$

The next set defines a relative potential depth between the group and aCCH₂, allowing alkyl CH₂ and CH₃ groups to be considered uniquely:

$$\varepsilon_{\text{rel},i} = \frac{\varepsilon_{i,\text{aCCH}_2}}{\varepsilon_{\text{aCCH}_2,\text{aCCH}_2}} \quad (3.14)$$

$$\alpha_n^3 = \frac{\varepsilon_{\text{rel},i}}{n} \quad (3.15)$$

$$\alpha_n^4 = \frac{\varepsilon_{\text{rel},i}}{n^2} \quad (3.16)$$

$$\alpha_n^5 = \frac{\varepsilon_{\text{rel},i}}{\exp(n)} \quad (3.17)$$

4. Results & Discussion

4.1 Parameter Selection

When optimising each set of scenarios, as outlined in section 3.2, four different initialisations were used and compared. Observing the results when optimising ξ alone, experiments 'A' and 'D' produce wildly inaccurate viscosity predictions compared to the data, while experiments 'B' and 'C' produce near-identical lines with much closer accuracy to the NIST curve. The optimised value of ξ in each case seems independent from the set of 'n' coefficients used, with scenario 'A' and 'B' generating a ξ_{opt} of

0.8527 whereas 'C' and 'D' produce a ξ_{opt} of 1.0246, in essence swapping their initialised ξ .

Table 4.1 Scenario parameter values post-optimisation; optimised values in bold

Opt.	ξ	n1	n2	n3
$\xi - A$	0.8527	0.252	-0.043	0.0004
$\xi - B$	0.8527	0.306	-0.128	0.0287
$\xi - C$	1.0246	0.252	-0.043	0.0004
$\xi - D$	1.0246	0.306	-0.128	0.0287
n - A	1.0	0.276	-0.043	0.0004
n - B	1.0	-3.843	3.563	-1.0785
n - C	0.8377	-2.794	2.329	-0.6568
n - D	0.8377	-2.794	2.329	-0.6568
$\xi + n - A$	0.2475	-0.311	0.125	-0.0216
$\xi + n - B$	0.3076	-0.461	0.210	-0.0397
$\xi + n - C$	0.2461	-0.308	0.123	-0.0213
$\xi + n - D$	0.3054	-0.455	0.207	-0.0389

The two successful models, post-optimisation, produce a set of parameters (ξ , n_x) close to the pairings used by Bell-Yang⁶. 'A' and 'D' are initialised with such pairings, but each change their ξ to produce either overly-specific or overly-generalised expressions that deviate from their intended application. The efficacy of the optimised parameter sets are as expected and align with previously recorded values in literature.

In optimising the 'n' coefficients, 'B', 'C' and 'D' produce near-identical outputs for viscosity. Only initialisation 'A' produces unique results - using values of 'n' that match those recorded by Bell & Yang (2025)⁶ the most closely - but producing the worst prediction in viscosity. Regardless of their uniqueness, it appears that optimisation with any respect to the 'n' parameters produces a concave curve, non-reflective of the convexity shown by viscosity with temperature. This may arise from significant adjustments being made to minimise errors around the vapour-liquid transition ($T \sim 350\text{K}$), rather than focusing only on the liquid phase.

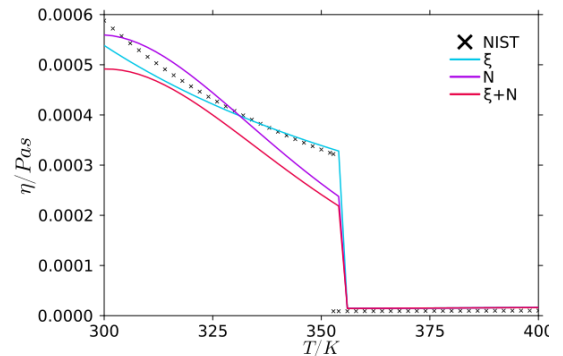


Figure 4.1 Comparison of the best initialisations of each optimisation scheme for benzene

The final scenario tested adjusts all parameters simultaneously. This suffered from similar failures

to that of ‘n’ alone, and none of these initialisations were particularly accurate in their modelling of the data. While possible that these issues are a failure of the optimisation method used, rather than the choice of parameter (hypothetically, more degrees of freedom should produce stronger results), for the sake of our work an efficient fitting of parameters was required for batch-training. If a certain parameterisation fails to do this accurately, it ought to be discounted regardless.

Both multi-parameter optimisations result in curves that fail to match the overall shape of the viscosity and, when comparing the best-cases of each scenario, it becomes apparent that optimising ξ alone is sufficient to producing strong correlations between the model and experimental data. Selecting ξ as the subject of our work also minimises risk of over-parameterisation, which may allow sets of data to be fitted arbitrarily rather than due to meaningful trends.

4.2 Overarching ξ Models

In the batch training sets (see Table (4.3)), all entropy-scaling methods provide reasonably strong results with mean Average Absolute Deviations (%) (AAD%) of less than 5. The modeling accuracy on the training set data represents a best-case scenario for each of these methods, and the predictive ability remains paramount in determining each one’s effectiveness.

Table 4.2 ξ_i Values for entropy-scaling methods

	Syst.	Sim.	Sim. T	VF	VF T
aCH	0.144	0.144	0.142	0.857	0.848
aCCH3	0.215	0.220	0.224	1.132	1.143
aCCH2	-0.178	-0.168	-0.158	1.929	1.998
CH2	0.038	0.034	0.037	1.344	1.375
CH3	0.409	0.422	0.413	0.464	0.401
cCH2	0.124	0.124	0.125	-	-
cCH	0.047*n - 0.194	0.047*n - 0.194	0.047*n - 0.194	-	-

The systematic approach finds itself in a unique situation, using a narrowed training set of only benzene, toluene, ethyl-benzene, hexane and octane, making comparisons to the batch-trained methods difficult aside from the molecules unseen to them all. For this reason, the systematic approach has been compared with other methods as though it were trained on the larger data set.

The systematic method performs extremely well for the molecules that it was trained on, having no incentive to alter its ξ_i values to minimise the errors of other molecules sharing the same groups. Consequently, it produces the smallest AAD% of the three entropy-scaling methods for the aforementioned species, and continues to perform well for n-alkanes. However, the increased accuracy for these specific molecules is outweighed by its failures in more complex molecules. For example,

errors for butyl-benzene, a molecule similar to ethylbenzene (that the systematic approach was trained on), are more than double those for the batch and ‘VolFrac’ methods. This arises as the method fails to accommodate for differences in how groups contribute differently within different molecules compared to its batch-trained counterparts.

Table 4.3 AAD% of Different Models on batch training set (prefixes with benzene)

Training Set	Sys	Sim.	Sim + T	VF	VF + T	Mol Dens
Benzene	2.27	2.29	3.08	2.41	2.69	7.40
Toluene	2.80	3.11	4.78	3.07	3.70	7.90
P-Xylene	3.62	0.79	2.39	1.48	0.95	9.32
Ethyl-	4.14	5.43	4.47	4.33	3.86	100.00
Propyl-	4.36	2.21	0.56	2.77	0.90	0.68
Butyl-	9.27	2.21	1.07	1.14	3.95	5.86
Pentyl-	13.27	4.36	0.55	2.32	3.14	11.70
Hexane	3.39	3.45	3.17	4.06	5.20	20.63
Heptane	1.70	2.06	3.66	2.66	1.79	12.28
Octane	1.69	1.89	2.29	4.50	3.45	11.21
Nonane	5.12	5.12	2.36	5.48	3.27	13.11
Decane	4.38	4.31	2.01	4.80	2.87	11.60
Dodecane	6.93	7.23	4.91	6.39	1.92	11.50
Mean	4.84	3.43	2.71	3.49	2.90	17.17
$\sqrt{\text{VAR}}$	3.16	1.72	1.42	1.51	1.19	24.31

Table 4.4 AAD% of Different Models on batch testing set (prefixes with benzene)

Testing Set	Systematic	Simple Batch	Simple + Temp
Trimethyl-	11.11	6.86	8.06
Hexyl-	4.38	12.11	17.13
Heptyl-	18.67	7.58	0.87
Octyl-	5.23	6.86	15.58
Dodecyl-	28.90	12.14	9.15
Mean	13.66	9.11	10.16
$\sqrt{\text{VAR}}$	9.18	2.48	5.83

Table 4.4 (cont.) AAD% of Different Models on batch testing set (prefixes with benzene)

Testing Set	Vol Frac	Vol Frac + Temp	Mol Dens
Trimethyl-	8.59	8.24	21.94
Hexyl-	10.11	16.09	12.24
Heptyl-	18.38	8.10	3.43
Octyl-	12.56	4.73	18.73
Dodecyl-	177.45	124.65	100.00
Mean	45.42	32.36	31.268
$\sqrt{\text{VAR}}$	66.10	46.29	34.94

Using the same model type with a batch-trained approach, named ‘simple batch’ - ‘Sim.’ in the tables - produces much stronger generalised results than systematic, with a mean AAD% of 3.43 and a standard deviation of 1.72. The molecules that the systematic approach has the upper-hand on, due to its highly specific ξ_i values, are modeled with negligible detriment to their accuracy, while drastic improvements are made on a number of other molecules, especially alkyl-benzenes.

Where the other two models mentioned thus far struggle with the largest of the training set molecules (see pentyl-benzene, dodecane in table 4.3), the ‘VolFrac’ model begins to prove some degree of applicability. For butyl-benzene and pentyl-benzene, significant improvements are made (48.4% and 46.8% better, respectively), and minor improvement is achieved for dodecane (11.6%), but other molecules appear to suffer for these improvements; namely octane, whose AAD% increased by a factor of 238.1%. With a slightly higher mean AAD% than Simple Batch, but slightly lower standard deviation, judgement of the models’ success is largely dependent on one’s preference for reliability or accuracy. It may again be noted that the form of ξ_{total} in this model is generally more agreeable with known molecular trends, in that it produces greater viscosity for larger molecules, and could prove more suitable with some adjustments.

Table 4.5 ξ_i Values for molar density-scaling method

	MolDens-n	MolDens-T
aCH	2.468	1.369
aCCH3	1.629	-0.700
aCCH2	3.479	-0.479
CH2	3.803	1.497
CH3	-1.104	-0.461
cCH2	-	-
cCH	-	-

Results for the molar density-scaling approach were indisputably poor compared to entropy-scaling, producing a mean AAD% of 17.17 and a standard deviation of 24.31 in the training set, despite promising trends being displayed in figures 3.5 – 3.7. Interestingly, the model estimates the viscosity of propyl-benzene to one of the highest accuracies displayed by any model across all molecules. Whether this is by chance, or a show that such a model could prove successful with some adjustment, is indeterminate without further exploration. Some of the failures of this model may be attributed to the abnormal progression of ideal y-intercepts for certain molecules in figure 3.7, like hexyl-benzene, whose jump from pentyl-benzene and then to heptyl-benzene is hugely inconsistent. Koenhen et al.² suggest that such deviations from typical entropic behaviour arise from chain-lengths beyond a certain point becoming more flexible, and thus capable of fitting with one another more easily, increasing viscosity. Producing models that take account for complex shapes and bonding between molecules are extremely challenging to develop in a generalised, simple-to-apply approach as is favoured in this work, and would require more advanced modelling techniques to produce consistent results. Furthermore, this approach introduces a secondary parameter for each group – the temperature gradient contribution – without providing any meaningful improvement to alternative models. In its current state, the density-scaled approach taken is

considered broadly unsuccessful and is discarded in favour of entropy-scaling.

Upon application to a set of five unseen molecules (table 4.4), all exhibiting aromaticity, the simple batch method proves both most accurate and most consistent in its prediction. Here, the systematic approach begins to falter in its prediction of heptyl-benzene and dodecyl-benzene, while the ‘VolFrac’ model hugely underestimates the required value of ξ for the larger molecules, such as dodecyl-benzene, suggesting the scaling used is too extreme, when unable to be compensated for, like in the training set. If considering its own testing set of molecules – containing all molecules but benzene, toluene, ethyl-benzene, hexane and octane – the systematic approach produces a mean AAD% of 8.99 and a standard deviation of 7.30. Its lower mean AAD% can be attributed to a greater inclusion of n-alkanes as testing molecules, whose group behaviours vary very little, making the systematic ξ values more generally applicable. The effect of alkyl chain lengths on viscosities for alkyl-benzenes displays more complex behaviour, and the systematic approach treats these groups as though they were aliphatic only, producing highly variable results. Further, the datasets used are relatively small in size, and a more developed training/testing range would serve to further improve the results of the batch methods compared to the systematic. As a result, the simple batch method is deduced to be the strongest model of the four tested thus far.

4.3 Adding Temperature Dependence

Applying the generalised temperature dependence factor as outlined in equation (3.10) typically results in a stronger performance for training datasets, but inconsistent applicability to predictive viscosities.

When applied to the simple linear combination, an improvement is made in both the mean and standard deviation of the training set AAD% compared to the temperature-independent function. Its predictive ability, in contrast, worsens in both volatility and general accuracy. No particular trend in its behaviour can be observed for the unseen set, with some molecules seeing dramatically improved predictions (i.e. Heptyl-benzene producing an AAD% of only 0.87) while others double their degree of inaccuracy, i.e. octyl-benzene. The two molecules most severely affected in either direction differ only in one CH2 group, so, while one may expect similar results between the two, this is not the case. Another strange outcome of this approach is that dodecane – a notable outlier in figure 3.10, deviating from the $y=x$ line – also sees improvement using the temperature dependent scaling, where some n-alkanes that agree more strongly with the $y=x$ line see worsened results, like in the case of octane. Ultimately, our model can only do so well as

to fit the most appropriate values of ξ at any viscosity point, but, even with ‘perfect’ values, there is a chance that the form of the Bell-Yang model has other flaws that result in inaccuracy regardless of ξ , making our adjustments somewhat unpredictable.

VolFrac + T represents the most complex of the entropy-scaling models used, and produces stronger results than VolFrac alone. All but hexyl-benzene - which appears to be a particularly anomalous molecule - benefit in the predictive testing set, dampening some of the failures of the VolFrac model. Dodecyl-benzene maintains a highly inaccurate prediction with an AAD% of 124.65 but, even with this improvement, remains highly inaccurate. The training set is also, generally, improved. It may be inappropriate to assume all molecules exhibit the same temperature relationship, and, as seen in the inconsistent application of temperature scaling, the variable sensitivity of different molecules depending on their size and groups seems to be an important factor to consider. The approach taken with the molar density scaling – introducing a secondary temperature gradient parameter for each group – appears to show some merit based on figures 3.5 – 3.7. However, this has only been applied to a density scaling method which may, itself, have flaws. Further testing of entropy scaling using a group-contribution temperature-dependent function should be investigated further, as it seems to show promise.

4.4 Chain Length Dependent Functions

Method 1, described in section (3.3), was tested on a wide range of naphthenes. These include n-alkyl-cyclohexanes ranging from C1 (methyl) to C10 (decyl). The optimised relationship for ξ_{cCH} was developed using C2, C4, C6 and C8 as testing data (denoted by the dashed lines in figure 4.2 and was found to follow a linear trend

$$\xi_{cCH} = An + B \quad (4.1)$$

Here A and B are constants to be fitted, optimised as 0.0471 and -0.194 respectively, and n denotes alkyl chain length. As shown in figure 4.2, performance was relatively strong, with all AAD% less than 10, and a mean AAD% of 6.40. No significant difference between the AAD% of the trained and untrained datasets (6.33 vs 6.47) was found; expected with the more simple nature of n-alkanes.

These results, across a wide range of temperatures and chain lengths, show initial promise. However, closer inspection reveals a minor mathematical inconsistency. A linear dependence of ξ_{cCH} on chain length is mathematically equivalent to defining a new ξ_{CH2} for use in n-alkyl-cyclohexanes, and implies that the alkyl chain exhibits different behaviour in a cyclic compound than in an n-alkane,

potentially due to the steric differences between the two molecular geometries. Nevertheless, the method appears robust, accurately predicting viscosities for unseen chain lengths both above and below the training range.

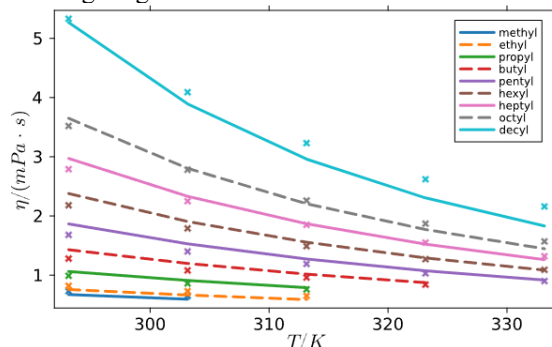


Figure 4.2 Viscosities of C1-C10 alkyl cyclohexanes where x marks experimental data¹⁰

Adding a simple temperature-dependent term, such as that seen in section 4.3, significantly improves predictions for C4 rings and smaller. The improvement is clear in figure 4.3 with the method yielding an AAD% of 3.18 for C4 and below. However, for larger naphthenes, this temperature correction scales poorly, leading to substantial underprediction of viscosity.

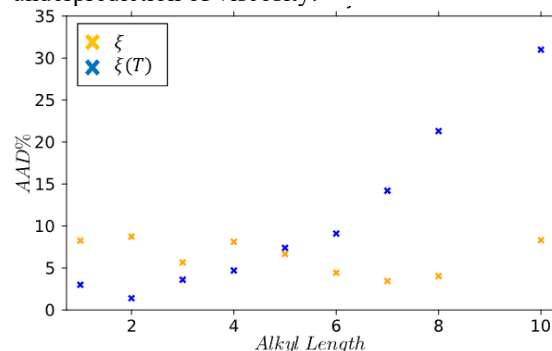


Figure 4.3 AAD% of alkyl cyclo-hexanes using the chain length method outlined in section 3.3

Using temperature dependence, the mean AAD% for C5 and above rises to 16.6, peaking at an AAD% of 31.1 for C10, suggesting that larger molecules or longer alkyl chains exhibit a stronger sensitivity to temperature. The relatively high melting points of these larger molecules could be the cause of this behaviour as, closer to the melting point, the gradient of viscosity with temperature becomes much steeper.

The addition of a chain-length dependent aCCH2 group showed no sign of improvement for any of the molecules considered. The most generalised forms of α produce greatly inaccurate results, and the inclusion of the scaling factor ε_{rel} helps to return to AAD% closer to that of $\alpha=0$. As only aCCH2 was optimised here (allowing use of CH2 and CH3 from n-alkanes), the inclusion of a relative depth for CH2 and CH3 serves to treat each group differently in

how significantly it affects the value of αCCH_2 , producing more specific and accurate results.

Table 4.6 AAD% for aromatic chain-dependent functions

	Ethyl-	Propyl-	Butyl-	Mean
$\alpha = 0$	4.12	0.76	2.13	3.28
α_1	16.37	5.84	11.61	11.27
α_2	11.52	5.36	7.29	8.06
α_3	6.57	2.37	2.59	3.84
α_4	6.13	2.30	2.40	10.83
α_5	4.86	1.50	1.38	2.58

However, it remains the case that the most simple linear combination ξ_i , without chain dependence, produces the strongest results. By adjusting αCCH_2 alone, sufficient accommodation for the occlusion from the n -alkyl chain seems to have been achieved; the degree that αCCH_2 is occluded appears to vary unsubstantially with chain length.

5. Conclusions

In this work, we present a new method that utilises group contributions to predict the viscosity of long-chain alkyl aromatics and naphthenes. The method employs SAFT- γ Mie to obtain the molecular entropy, followed by a modified Bell and Yang entropy-scaling approach. Because the method is based on group contributions, no pure-component parameters are required. This enables a predictive framework that maintains accuracy across a wide range of aromatics and naphthenes. The resulting model is robust and scalable, providing broadly accurate predictions. Incorporating temperature dependence can selectively improve accuracy for most molecules.

The behaviour of ξ for larger molecules exhibits inherent limitations, as ξ increases indefinitely, leading to underestimated viscosities. The model may also perform poorly near the pour point, where the temperature dependence of viscosity deviates markedly from higher-temperature behaviour. Further refinement of the temperature-dependent formulation should be investigated. In addition, this work has focused predominantly on viscosity data at atmospheric pressure, introducing pressure dependence into the n_1 , n_2 , and n_3 parameters. Finally, preliminary work on mixture behaviour did not reveal a simple solution. Further exploration of methods that either combine ξ or directly combine viscosities is therefore recommended.

7. Acknowledgements

We would like to thank Harrison Fraser, Eric Huang and Daanyaal Sajed for their invaluable assistance over the course of this work.

8. References

- (1) Burnett, Nuala, and Iona Stewart. "The UK's Plans and Progress to Reach Net Zero by 2050." *House of Commons Library*, UK Parliament, 26 Sept 2024, <https://commonslibrary.parliament.uk/research-briefings/cbp-9888/>
- (2) U.S. Department of Energy. "Alternative Fuels Data Center: Sustainable Aviation Fuel." <https://afdc.energy.gov/fuels/sustainable-aviation-fuel>
- (3) Michaga, Maria Fernanda Rojas, et al. "Techno-Economic and Life Cycle Assessment Review of Sustainable Aviation Fuel Produced via Biomass Gasification." *Sustainable Biofuels*, 2021, pp. 269–303, <https://doi.org/10.1016/b978-0-12-820297-5.00012-8>.
- (4) Lötgering-Lin, Oliver, and Joachim Groß. "Group Contribution Method for Viscosities Based on Entropy Scaling Using the Perturbed-Chain Polar Statistical Associating Fluid Theory." *Industrial & Engineering Chemistry Research*, vol. 54, no. 32, 6 Aug. 2015, pp. 7942–7952, <https://doi.org/10.1021/acs.iecr.5b01698>
- (5) Chapman, Sydney, et al. "The Mathematical Theory of Non-Uniform Gases." *American Journal of Physics*, vol. 30, no. 5, May 1962, pp. 389–389, <https://doi.org/10.1119/1.1942035>.
- (6) Martinek, Viktor, et al. "Entropy Scaling of Viscosity IV—Application to 124 Industrially Important Fluids." *Journal of Chemical & Engineering Data*, vol. 70, no. 2, 10 Jan. 2025, pp. 727–742, <https://doi.org/10.1021/acs.jced.4c00451>
- (7) National Institute of Standards and Technology. "Thermophysical Properties of Fluid Systems." *NIST*, 2018, webbook.nist.gov/chemistry/fluid/.
- (8) Luning, Dianne J, et al. "Densities and Viscosities at 293.15–373.15 K, Speeds of Sound and Bulk Moduli at 293.15–333.15 K, Surface Tensions, and Flash Points of Binary Mixtures of N -Hexadecane and Alkylbenzenes at 0.1 MPa." *Journal of Chemical & Engineering Data*, vol. 62, no. 5, 26 Apr. 2017, pp. 1673–1688, <https://doi.org/10.1021/acs.jced.7b00087>.
- (9) Koenhen, D.M et al. (1984). Phase separation phenomena in solutions of poly(2,6-dimethyl-1,4-phenylene oxide). IV. Thermodynamic parameters for solutions in a series of homologous solvents: Toluene to hexylbenzene. *Journal of Polymer Science: Polymer Physics Edition*, 22(12), pp.2145–2157. <https://doi.org/10.1002/pol.1984.180221212>
- (10) Jeanne, Dianne. " N -Pentadecane and N -Alkylcyclohexane Mixtures: Viscosity Deviations and Excess Molar Volumes, Speeds of Sound, and Isentropic Compressibilities at 0.1 MPa." *Journal of Chemical & Engineering Data*, vol. 68, no. 9, 10 Aug. 2023, pp. 2198–2211, <https://doi.org/10.1021/acs.jced.3c00290>.

Optimisation and Evaluation of a Lipstick-Based Colorimetric Biosensor for Salivary pH Analysis Using Red Cabbage Powder

Alexa Chew (02192714), Bam Tewinpagti (02203199)

Department of Chemical Engineering, Imperial College London, U.K.

Abstract Wearable biosensors offer a promising route for accessible, non-invasive health monitoring. This study explored a lipstick-based colorimetric biosensor with red cabbage powder (RCP) as a natural pH indicator for salivary pH detection. Lipstick formulations were varied in beeswax, mineral oil, and RCP content to optimise machinal stability, smooth application, and colour performance. Qualitative assessments and rheological characterisation were performed to evaluate hardness and texture, while absorbance and reflectance measurements were taken to analyse the biosensor's colorimetric response when incorporated into lipstick formulation. The optimal formulation, consisting of 1.5 g beeswax, 1.0 g mineral oil, and 1.0 g RCP, achieved an ideal balance between cosmetic properties and biosensor functionality. Optical results showed that anthocyanin remained chemically active within the wax-oil matrix, producing measurable wavelength shift and visible colour change with increasing pH. These findings confirm the feasibility of a lipstick-based pH biosensor and highlights opportunities to enhance pigment accessibility for stronger colorimetric response.

1. Introduction

1.1 Biosensors

Biosensors work by recognising biological or chemical reactions and producing signals that indicate how much of a particular analyte is present. They contain three key parts: a sensing element that detects the analyte, a transducer that converts this detection into a signal, and a detector that records and interprets the response [1]. Because of their sensitivity, selectivity, and rapid response, biosensors are increasingly used in fields such as medical diagnostics, food quality control, environmental monitoring and drug discovery [2]. In the medical field, biosensors have revolutionised disease diagnosis and management. Glucose biosensors, for instance, are widely used for diabetes monitoring. Fluorescent biosensors are also used in cancer research and drug discovery, providing valuable insights into the role and regulation of enzymes at the cellular level [3].

1.2 Limitations of Current Biosensing Technologies

Advances in wearable technologies have enabled continuous and non-invasive monitoring of biomarkers in bodily fluids (i.e. sweat, saliva, tears). These can take on various forms, including smart watches, eyeglasses, tattoos, gloves, clothing, or even implants [3]. Despite their promise, existing wearable biosensors present several limitations. Sweat-based patches often require frequent calibration to maintain accuracy and may cause skin irritation or contact dermatitis [4]. Tattoos and transdermal sensors, while innovative in concept, also present practical difficulties. They often struggle to maintain stable adhesion, particularly in humid or moist environments and may need replacement once they become wet. This restricts their durability and makes many of these devices suitable for only a single use. These highlight the need for more comfortable, intuitive, and user-

friendly biosensing platforms that fit naturally into everyday routines.

1.3 Saliva as a Non-Invasive Diagnostic Fluid

Saliva is a complex biological fluid secreted by major and minor salivary glands, serving key physiological roles in lubrication, digestion, buffering, and antimicrobial protection [5]. Importantly, saliva contains diverse biomarkers such as enzymes, hormones, antibodies, and metabolites that reflect oral and systemic health. Saliva is an attractive alternative to blood and sweat for non-invasive biosensing as it can be collected easily, painlessly, and without the need for specialist equipment. Its accessibility and low cost make it particularly suitable for point-of-care diagnostics.

However, current saliva biosensors are limited in practicality. Most detect only a few biomarkers or require deliberate sample collection with test strips or handheld devices [6], while wearable sensors like mouthguards are often uncomfortable for prolonged use. These limitations highlight the need for a biosensing platform that integrates seamlessly into daily routines without additional equipment.

1.4 Lipstick as an Ideal Biosensor Platform

Lipstick offers a promising platform for colorimetric biosensor because it is applied directly to the lips and naturally encounters saliva during everyday activity. This allows color change to occur passively, without the need for additional device or behavioral changes. Its routine use, portability, and visual nature further support its suitability as a sensing medium. Incorporating a pH-responsive pigment into lipstick therefore creates an opportunity for discreet, user-friendly biosensing that fit with everyday use.

1.5 Red Cabbage Powder as a pH Indicator

Red cabbage powder is rich in anthocyanins, water-soluble pigments that undergo pH-dependent structural transformations as hydrogen ion concentration varies, shifting the chromophore and producing colours from red to purple, blue, and green. At very acidic pH, the dominant species is the flavylium cation, which appears red. As pH increases, the pigment converts to the quinoidal base, generating purple-blue hues. Under near neutral conditions, anthocyanins primarily exist in the carbinol (pseudobase) form, resulting in a colorless appearance. In alkaline conditions, the molecule undergoes ring opening to form chalcone, producing green to yellow tones. These structural interconversions, driven by hyperchromic and bathochromic shifts, generate a clear stepwise color progression across the pH spectrum, as shown in Figure 1.5 [7].

Red cabbage anthocyanins are inexpensive, naturally abundant, non-toxic, and strongly halochromic, making them attractive for use in colorimetric indications and pH-responsive materials. However, incorporation into a hydrophobic wax-oil matrix in lipstick fabrication poses challenges, as pigment activity must be preserved to ensure reliable pH-responsiveness.

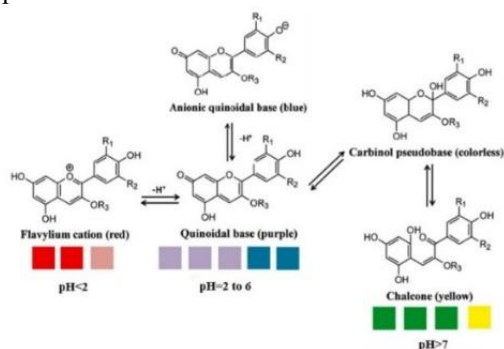


Figure 1.5 Structural interconversion of anthocyanin species across different pH conditions [7]

1.6 Academic Work

The wearable biosensor field is a popular and growing research area as the healthcare sector demands more accessible and non-invasive health monitoring technologies. However, the application of biosensors in lip-based cosmetics for health monitoring has yet to be researched extensively, presenting a strong potential for further study in the subject area. There is currently only one academic study conducted on lip-based biosensors: BioCosMe, a lip-based cosmetic with red cabbage powder acting as a colorimetric biosensor [8]. The study introduced three lip-based products – lipstick, lip gloss, and lip tint – containing red cabbage powder as the biosensor to detect salivary

pH levels. The study encompassed three main sections:

1. Formulation development for the three form factors, with technical evaluation of reversibility and re-usability.
2. Development of a Convolutional Neural Network (CNN) model for detection of pH levels on lips via taking photographs.
3. User study to understand usability of BioCosMe.

However, the product is still in the early stages of development and further study is required to validate the reproducibility of lip formulations developed and results obtained, and quantitatively determine the accuracy and efficacy of the developed lipstick biosensor.

1.7 Aim of This Study

This study sought to build on the BioCosMe study by focusing on formulation optimisation and evaluation utilising multiple experimental techniques to evaluate hardness, texture, and colorimetric biosensor functionality. This study focused only on lipstick formulation as it was deemed the most attractive form factor out of the three investigated. The specific objectives were to:

1. Optimise the lipstick formulation by varying beeswax, mineral oil, and red cabbage powder (RCP) content to achieve suitable mechanical and cosmetic properties.
2. Assess the colorimetric responsiveness of the lipstick across relevant pH conditions (6.0 – 8.0).
3. Explore alternative optical measurement techniques to determine the most accurate method for analysing lipstick colorimetric response.
4. Evaluate the feasibility of a pH-responsive lipstick as a practical biosensor and identify limitations and future design improvements.

2. Materials and Methods

2.1 Lipstick-Based Biosensor Ingredient Selection

Firstly, the relevant ingredients were selected to be used in proceeding lipstick formulation development experiments. The aim was to keep the lipstick formulation as simple as possible (i.e. with few ingredients) to avoid any adverse effects on biosensor colorimetric function, while still achieving a formula and feel that was similar to commercial lipstick formulations.

Waxes (10–25%)
Butters (10–15%)
Oils (30–50%)
Texturizers and Fillers (8–12%)
Pigments and Pearls (5–20%)
Antioxidants and Preservatives (around 1%)
Parfum (1–2%)
Special Additives (1–5%)

Figure 2.1 Average composition of a lipstick formula [9]

Figure 2.1 shows the average composition of a lipstick formula. Waxes and oils were determined to be essential to lipstick formulation; waxes are essential to provide structure, hardness, and shape retention to a lipstick, while oils provide lubrication, gloss, and comfort on application. Beeswax was selected due to its moderate hardness and smooth texture and has emollient properties that retain moisture on the lips. Mineral oil was selected as it is inexpensive and commonly used in cosmetics as an emollient and moisturiser. The pigment of choice in this case would be our biosensor, red cabbage powder, containing anthocyanin (see Section 1.5). All other components in Figure 2.1 were deemed secondary to essential lipstick function and were not utilised to avoid adverse effects on biosensor functionality.

2.2 Lipstick Fabrication Process

The lipstick samples were prepared using a standardised method for all experiments. First, all ingredients were weighed according to the formulation requirements. Silicone moulds were pre-chilled in the refrigerator for at least 30 minutes to promote rapid setting. Beeswax was melted at 75°C for 15 minutes with continuous stirring. Once fully melted, the remaining ingredients – red cabbage powder and mineral oil – were added to the beaker, and the mixture was stirred for an additional 10 minutes. The molten lipstick formula was then poured into the moulds and placed in the refrigerator for 15 minutes to solidify before demoulding. The lipstick sample is then stored at room temperature till usage.

2.3 Lipstick Formula Optimisation

A one factor at a time strategy was utilised to investigate how each formulation component – mineral oil, beeswax, and red cabbage powder (RCP) - influenced the physical properties of lipstick, such as hardness, texture, and colour payoff, with the goal of achieving lipstick properties comparable to commercial lipstick formulation. In each experiment, one variable was varied while all other components were kept constant. The experiments were done sequentially and obtained results for the optimum amount of each component were carried forward to the succeeding experiments. All detailed ingredient weights for each experiment are provided below in Table 2.3.1 to 2.3.3.

Table 2.3.1 Experiment 1: variation of mineral oil content

Mineral Oil (g)	Beeswax (g)	RCP (g)
0.0	1.5	1.0
1.0	1.5	1.0
2.0	1.5	1.0
3.0	1.5	1.0
4.0	1.5	1.0

Table 2.3.2 Experiment 2: variation of beeswax content

Mineral Oil (g)	Beeswax (g)	RCP (g)
X.X	0.0	1.0
X.X	0.5	1.0
X.X	1.0	1.0
X.X	1.5	1.0
X.X	2.0	1.0

Table 2.3.3 Experiment 3: variation of red cabbage powder (RCP) content

Mineral Oil (g)	Beeswax (g)	RCP (g)
X.X	X.X	0.0
X.X	X.X	0.5
X.X	X.X	1.0
X.X	X.X	1.5
X.X	X.X	2.0

Multiple studies were conducted in tandem that influenced the selection process. A qualitative study was conducted on observations of hardness and texture, and rheological measurements taken to verify the optimum lipstick formula selected. Optical measurements were then conducted to validate the pH-responsiveness of red cabbage powder when incorporated in lipstick formulation.

2.3.1 Qualitative Rating of Hardness and Texture

Qualitative rating scales were developed to define the hardness and texture characteristics of tested lipstick formulas. Hardness is defined as the resistance to deformation from force, while texture is the overall feeling from surface characteristics upon application, such as roughness or smoothness. In the final optimised lipstick formula, a balance of hardness and texture must be achieved to allow for ideal lipstick function; as if a lipstick is too hard, it is likely to have a rough texture upon application and would have significant resistance upon application. Table 2.3.1.1 and 2.3.1.2 below illustrate the qualitative ratings for hardness and texture respectively.

Table 2.3.1.1 Qualitative rating scale for hardness (1=Extra Soft, 6=Extra Firm)

Number Scale	Hardness	Qualitative Observation
1	Extra Soft	Lipstick does not hold shape of mould after cooling and removal. Forms a liquid-like paste.
2	Soft	Lipstick does not hold shape of mould after cooling and removal. Forms a clay-like paste.
3	Medium Soft	Lipstick adheres to shape of mould after cooling and removal. Shape deforms significantly upon light pressure from touch and application on a surface.
4	Medium Firm	Lipstick adheres to shape of mould after cooling and removal. Shape deforms slightly upon light pressure from touch and application on a surface.
5	Firm	Lipstick adheres to shape of mould after cooling and removal. Shape does not deform upon light pressure from touch but deforms slightly upon pressure from application on a surface.
6	Extra Firm	Lipstick adheres to shape of mould after cooling and removal. Shape does not deform upon light pressure from and application on a surface.

Table 2.3.1.2 Qualitative rating scale for texture (1=Extra Rough, 6=Extra Smooth)

Number Scale	Texture	Qualitative Observation
1	Extra Rough	Lipstick applies with significant resistance and forms a non-uniform layer with significant streaking and patchiness.
2	Rough	Lipstick applies with noticeable resistance and forms a non-uniform layer with slight streaking and patchiness.
3	Medium Rough	Lipstick applies with moderate resistance and requires moderate pressure and effort to form a mostly uniform layer.
4	Medium Smooth	Lipstick applies with moderate resistance and requires light pressure and effort to form a uniform layer.
5	Smooth	Lipstick applies with minimal resistance and forms a uniform layer.
6	Extra Smooth	Lipstick applies with no resistance and forms a uniform layer.

2.3.2 Rheological Characterisation

The aim of this study was to characterise the mechanical properties of the tested lipstick formulas to verify the qualitative hardness rating and observations from Section 2.3.1 with quantitative rheological data. Furthermore, this would allow this study to confirm the hypothesis that with an increased beeswax content (and conversely a decreased oil content), the lipstick would be “harder” and more resistant to deformation upon touch and application.

The Anton Paar Modular Compact Rheometer (MCR) 302 was utilised for all measurements, fitted with a 25mm parallel-plate, stainless steel geometry and utilised a 1mm gap. The lower plate is connected to a Peltier system, which is further allied to a water bath to regulate the temperature of the plate within $\pm 0.01^{\circ}\text{C}$. It was critical that a constant volume and geometry of the sample is maintained for all readings to obtain a smooth oscillatory torque signal. Each lipstick sample was freshly excised with a spatula and flattened into the thinnest possible layer, then placed on the lower plate as the upper plate lowers down to a sample height of 1mm. Once this thickness was reached, excess material was removed with a right-angle spatula, and the sample was left for 15 minutes to set before commencing measurement.

Isothermal amplitude sweeps were conducted at 25°C for each sample to measure the linear viscoelastic region (LVER), obtaining measurements of storage modulus G' , loss modulus G'' , and shear stress τ across an oscillatory shear strain range of $0.001\% < \gamma < 100\%$ (log scale) at a constant angular frequency of $\omega = 1 \text{ rad s}^{-1}$.

2.3.3 Absorbance Measurement

This study sought to measure the absorbance spectrum of the lipstick samples tested to determine if varying contents of mineral oil, beeswax, and red cabbage powder influenced the colorimetric response of the overall lipstick formula. Additionally, the baseline colorimetric response of red cabbage powder in buffer pH solutions was established. This was done to observe if the lipstick form factor had any effect on biosensor response by allowing comparison of the biosensor response in solution versus incorporated in the lipstick formula.

2.3.3.1 pH Buffer and Absorbance Equipment Setup

pH buffer solutions of pH 6.0 to 8.0 in 0.5 increments were utilised across all testing. This is selected to accurately represent the natural human pH saliva range of 6.2 to 7.6 [10]. The pH buffer solutions were prepared using sodium phosphate monobasic (NaH_2PO_4 , MW 119.98 g/mol) and sodium phosphate dibasic (Na_2HPO_4 , MW 141.96 g/mol), which allows

for a pH buffer range of 5.8-8.0 at 25°C. Two bottles of 500ml deionised water were prepared, and 8.999g of NaH_2PO_4 and 10.647g of Na_2HPO_4 were dissolved in each bottle respectively for required buffer concentration of 150 mM. An InLab micro pH meter was then used to mix the two stock solutions to achieve pH buffer solutions at pH of 6.0, 6.5, 7.0, 7.5 and 8.0.

For all absorbance measurements, a Thermo Scientific™ Varioskan™ LUX Multimode Microplate Reader was used with standard 96-well plates. The absorbance of all samples was measured, where the microplate reader shines a light source vertically through each well at a specified wavelength range of 400 to 800 nm in the visible spectrum. The software then utilises the difference between the initial and transmitted light intensity to calculate the absorbance.

2.3.3.2 Absorbance Measurement of Red Cabbage Powder in Solution

Firstly, the baseline colorimetric response of red cabbage powder in solution was investigated. 10mg of red cabbage powder (RCP) was weighed and mixed with 2ml of each pH buffer solution of 6.0 to 8.0 in 0.5 increments to minimise material loss during transfer. The mixture was then transferred into a glass vial, and an additional 8 ml of buffer was added to obtain a total volume of 10 ml. From each RCP-buffer mixture, 100 μl was pipetted into individual wells on the 96-well plate, and absorbance spectra recorded from 400 to 800 nm in triplicate using the microplate reader. The wavelength at which the absorbance peak is found is then recorded and plotted against pH.

2.3.3.3 Absorbance Measurement of Solid Lipstick Sample

Each lipstick formulation sample tested is then investigated to determine if varying contents of mineral oil, beeswax, and red cabbage powder had an influence on the biosensor's colorimetric response. 10mg of each sample was placed in the individual wells of a 96-well plate and flattened with a glass rod to an even thickness, ensuring the sample covers the entire bottom of the well. 100 μl of pH buffer solution is then dropped into the well with a 20-200 μl pipette. Each lipstick sample is tested with pH solutions of 6.0 to 8.0 in 0.5 increments, and absorbance spectra recorded from 400 to 800 nm in triplicate using the microplate reader. The absorbance peak wavelength is then recorded and plotted against pH.

2.4 Alternative Absorbance Measurement with Spectrophotometry

Once the final optimum lipstick formulation was selected, an Ocean Optics Flame miniature

spectrometer was used to measure the reflectance spectrum of the optimum lipstick sample exposed to different pH buffer solutions from 6.0 to 8.0 at 0.5 increments; this allowed for comparison with the absorbance measurements done from the microplate reader via transmittance. This was done to determine which technique provided results of a greater accuracy and whether the results reflected visual observations of colorimetric response.

Lipstick was applied to a glass slide using standardise application protocol with controlled pressure, number of strokes, and surface area to ensure uniform layer. For each test, 400 μl of pH buffer solution was added to the swatch and left for 30 minutes before removing excessing liquid, allowing the biosensor's colorimetric response to develop fully. The slide is then clamped beneath the spectrometer's receiver at a fixed distance. The reflectance spectrum across the visible spectrum (400-800nm) was then recorded for each condition. The wavelength at which the absorbance peak is observed was extracted and plotted against pH.

2.5 Visual Biosensor Colorimetric Response

Finally, photos of all samples from Section 2.4 were taken under identical lighting conditions, setup, and camera angle using a phone stand. Photos were also taken of pH buffers with red cabbage powder dissolved in solution from Section 2.3.3.2 under the same conditions. Each photo was then cropped to a reference size kept constant for all samples where the colour was most concentrated. This was done for visual qualitative observation of the biosensor's colorimetric response for comparison to quantitative absorbance results obtained.

3. Results

3.1 Qualitative Analysis

3.1.1 Mineral Oil Variation

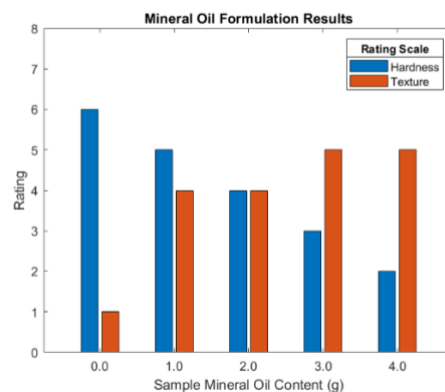


Figure 3.1.1 Hardness and texture ratings for lipstick formulations with varying mineral oil content

Figure 3.1.1 shows the effect of varying mineral oil content on the qualitatively observed properties of the lipstick. An increased oil concentration resulted in reduced hardness, with the highest hardness score observed at 0.0g oil and the lowest at 4.0g oil. Texture score followed the opposite trend: formulations with high oil concentration appeared smoother and easier to apply, whereas low-oil samples were less spreadable. Therefore, a mineral oil content of 1.0g was deemed to provide an optimal balance between texture and structural stability.

3.1.2 Beeswax Variation

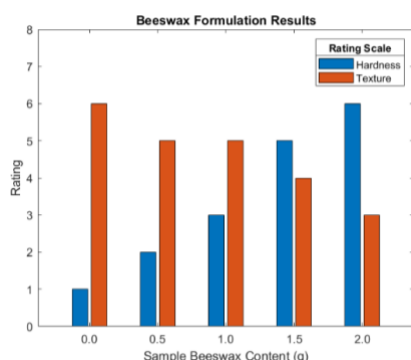


Figure 3.1.2 Hardness and texture ratings for lipstick formulations with varying beeswax content

Figure 3.1.2 shows the effect of varying beeswax content on the physical properties of the lipstick. Increasing beeswax concentrations resulted in an increased hardness. Samples with low beeswax content were too soft to demould and did not retain their shape. Texture scores followed the opposite trend: lipsticks with a low wax content were smooth and spreadable. Based on the combined scoring, 1.5g beeswax was chosen to provide the optimal balance between hardness and texture.

3.1.3 Red Cabbage Powder (RCP) Variation

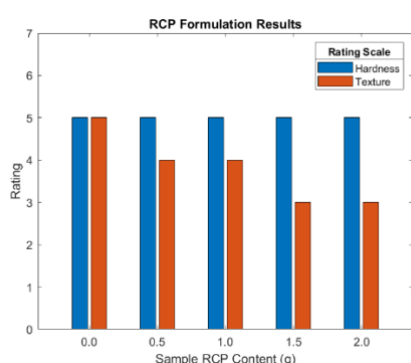


Figure 3.1.3 Hardness and texture ratings for lipstick formulations with varying red cabbage powder (RCP) content

Figure 3.1.3 shows the effect of varying red cabbage powder (RCP) content on the physical properties of the lipstick. The hardness rating for all samples remained constant as the oil and wax content was fixed. A RCP content of 0.0g was not chosen as it does

not incorporate the biosensor in the lipstick. Intermediate RCP contents of 0.5g and 1.0g showed the best performance in texture, while RCP contents higher than 1.0g were less optimal due to pigment overload leading to particle clustering. However, a red cabbage powder content 0.5g produced a lipstick with insufficient colour saturation to allow for appropriate biosensor functionality. Based on this and the combined scoring of hardness and texture, 1.0g RCP achieved the most consistent colour and biosensor intensity while maintaining a smooth surface. Finally, the final optimal formulation based on the qualitative study is shown in Table 3.1.

Table 3.1 Composition of the optimal lipstick formulation

Beeswax (g)	Mineral Oil (g)	Red Cabbage Powder (RCP) (g)
1.5	1.0	1.0

3.2 Rheological Characterisation

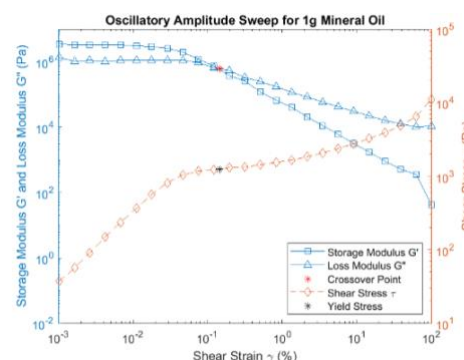


Figure 3.2.1 Oscillatory amplitude sweep for lipstick formulation with mineral oil content of 1.0 g

Figure 3.2.1 shows the full mechanical spectrum of the lipstick formulation with 1.0g as a function of applied strain, and all other oscillatory amplitude sweeps for other mineral oil contents can be found in Supplementary Information (Figure S.1). All samples are only viscoelastic up to a low strain of ~0.1% (see Figure 3.2.2) where $G' > G''$, and G'' is dominant beyond this point. This illustrates that all samples at different levels of oil content become more viscous and less elastic at higher strain, where the internal wax breaks down, undergoing structural collapse beyond this point.

Figure 3.2.3 illustrates a strong negative correlation between G' in the linear viscoelastic region (LVER) versus lipstick mineral oil content ($R^2=0.9761$). For all lipstick samples, the G' plateaus in the 0.001-0.1% strain range (see Figure A.1); three G' values were taken in the plateau region and the mean calculated. Additionally, Figure 3.2.4 indicates a weak negative correlation between yield stress and lipstick mineral oil content ($R^2=0.4710$). The yield stress at a mineral oil content of 1g is the highest of all samples at 1241 Pa. This suggests that the optimal wax-oil ratio is

achieved at 1g mineral oil and 1.5g beeswax, in line with the optimum formula determined from qualitative study in Section 3.1.

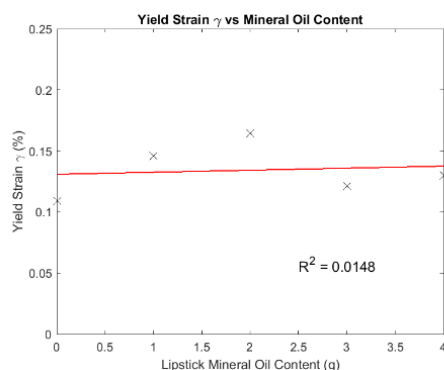


Figure 3.2.2 Yield strain γ (%) versus lipstick mineral oil content

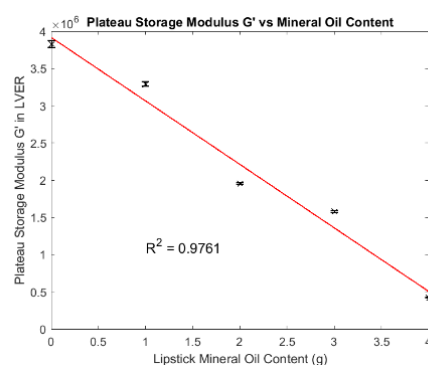


Figure 3.2.3 Plateau storage modulus G' in linear viscoelastic region (LVER) versus lipstick mineral oil content

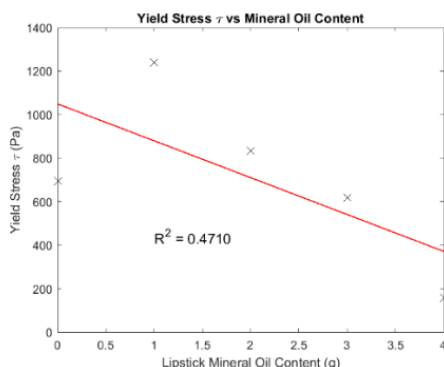


Figure 3.2.4 Yield stress τ (Pa) versus lipstick mineral oil content

3.3 Absorbance Analysis

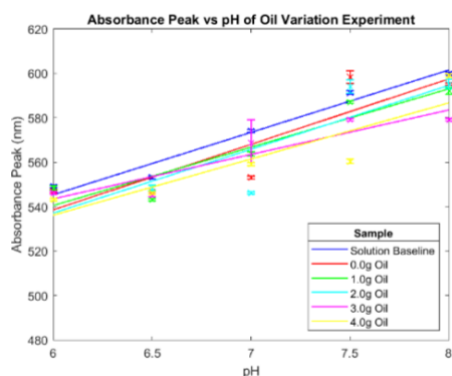


Figure 3.3.1 Absorbance peak vs pH for red cabbage powder (RCP) in solution and lipstick with varying mineral oil content

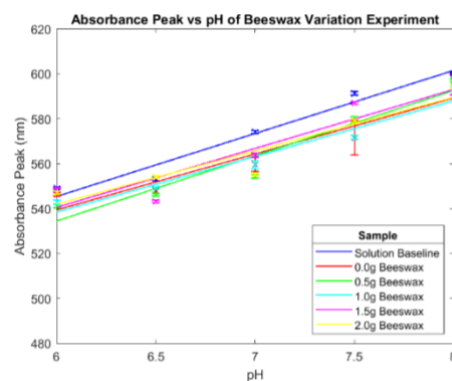


Figure 3.3.2 Absorbance peak vs pH for red cabbage powder (RCP) in solution and lipstick with varying beeswax content

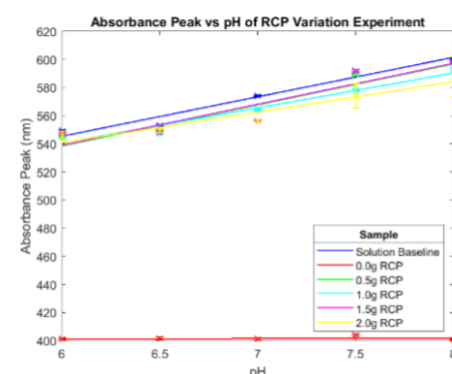


Figure 3.3.3 Absorbance peak vs pH for red cabbage powder (RCP) in solution and lipstick with varying RCP content

Table 3.3.1 Calculated R^2 values for each lipstick formulation tested and red cabbage powder in aqueous solution

Oil Content (g)	R^2 Value	Beeswax Content (g)	R^2 Value	RCP Content (g)	R^2 Value
0.0	0.7876	0.0	0.9041	0.0	0.0219
1.0	0.8799	0.5	0.9307	0.5	0.9023
2.0	0.7541	1.0	0.9504	1.0	0.9548
3.0	0.8653	1.5	0.8799	1.5	0.8670
4.0	0.7995	2.0	0.8969	2.0	0.9374
R^2 Value					
RCP in Solution (Baseline)				0.9649	

Figure 3.3.1 to 3.3.3 show the absorbance peak across pH 6-8 for lipstick formulations with varying oil, beeswax and RCP concentration, compared with the RCP solution baseline. All samples displayed a positive linear correlation with pH, consistent with the expected behaviour of anthocyanins. The solution baseline showed the strongest linearity ($R^2 = 0.96$), while lipstick formulations showed slightly weaker due to the solid. As expected, the lipstick formulation with 0.0g RCP did not show any trend with pH due to no pigment incorporated. To assess measurement reliability, 95% confidence intervals were calculated for all measurements. Across every formulation and experimental condition, all replicate values fell within their respective 95% CIs, indicating high repeatability and the absence of outliers in the dataset. These results demonstrate that RCP remained pH-responsive after being incorporated into the lipstick, confirming that the formulation process did not degrade the pigment.

However, the magnitude of the colour response detected by the microplate reader was substantially greater than the colour change observed visually on the lipstick surface (see Section 3.5).

3.4 Spectrophotometry Analysis

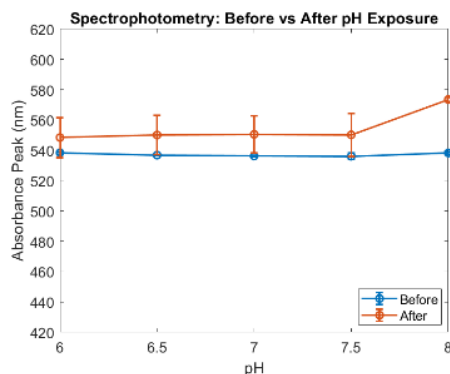


Figure 3.4 Peak wavelength of the optimal lipstick before and after pH exposure

Figure 3.4 illustrates that before pH exposure, the peak wavelength remained fairly constant across all conditions, ranging between 536 and 538 nm. After exposure to the pH buffer, the lipstick surface exhibited a clear upward shift in wavelength at every pH tested, rising to approximately 550 nm for pH 6.0-7.5, with a larger shift observed at pH 8.0. Although the pattern was not strictly linear, all after-treatment values were consistently higher than the untreated baseline, demonstrating that the lipstick surface underwent a measurable change in its absorbance peak following pH exposure.

3.5 Visual Biosensor Colorimetric Response

The swatch images in Table 3.5.1 show the appearance of the lipstick surface after removal of the pH buffer. A gradual darkening of the lipstick colour is observed with increasing pH, which aligns with the upward shift in peak wavelength and supports the spectrophotometry data. The colour change is subtle and appears as a deep purple rather than bright blue typically seen in solution, where Table 3.5.2 presents the corresponding colour transition for red cabbage powder in aqueous solution.

Table 3.5.1 Colour change of the lipstick after pH exposure

No pH	6	6.5	7	7.5	8

Table 3.5.2 Colour change of the RCP solution across pH levels

6	6.5	7	7.5	8

4. Discussion

4.1 Influence of Formulation Components on Mechanical Properties

The balance between mineral oil, beeswax, and red cabbage powder strongly governs the physical behaviour of the lipstick. Increasing mineral oil content softens the lipstick and reduces its resistance to deformation, as shown in Figure 3.2.3 by the decrease in G' within the linear viscoelastic region (LVER). A small amount of mineral oil enhances structural integrity by acting as a plasticiser, facilitating closer packing and cohesion within the beeswax crystalline network, producing a more continuous, stress-resistant matrix. This agrees with literature on oleogels, where moderate oil levels improve structural uniformity and elasticity, but excessive liquid phase dilutes crystal packing, resulting in lower hardness and yield stress [11].

Conversely, a higher beeswax content strengthens the internal structure, resulting in firmer, more rigid lipsticks with higher G' values in the LVER, reflecting a denser crystal network. However, excessive wax content reduces spreadability and application smoothness, highlighting a trade-off between mechanical stability and usability when selecting an optimal lipstick formulation. Red cabbage powder minimally affects mechanical strength, where the pigment particles function primarily as an inert filler within the matrix; at high loadings, pigment agglomeration reduces smoothness and produces a rougher lipstick texture.

Because of this relationship between formulation components, it can be argued that the relationship observed in Figure 3.2.4 between yield stress and mineral oil content is non-linear, showing a weak negative correlation ($R^2=0.4710$). Maximum yield stress of 1241 Pa occurs at 1g mineral oil content, where plasticisation optimises crystal packing. With no oil, the beeswax network becomes brittle and discontinuous, leading to poor packing and voids. Although G' is highest at zero oil content, the yield stress is low as cracks propagate easily under shear. Beyond a mineral oil content of 1g, further dilution of the beeswax matrix lowers yield stress.

Crossover points where $G'' > G'$ for all samples occur consistently at $\sim 0.1\%$ strain (Figure 3.2.2), suggesting similar failure mechanisms of the wax network across formulations. Mineral oil mainly scales the stress level at which failure occurs rather than the strain at which it begins. All samples exhibit narrow linear viscoelastic regions, yielding at $\sim 0.1\%$ strain, much lower than the 10-20% reported for commercial lipsticks [12]. This brittle response is likely due to the simplified lipstick formulation in this study to allow for biosensor integration and functionality, and the

absence of softer waxes or emollients (Section 2.1). Rapid cooling or incomplete melting during fabrication may also promote large wax crystal formation or microcracks, reducing yield strain. Nonetheless, this study's formulation remains suitable for cosmetic use, as shear during application onto the lips is minimal.

4.2 Performance Evaluation of Optical Measurement Techniques

The microplate reader produced absorbance spectra closely resembling the solution baseline rather than the lipstick's true colorimetric behaviour observed visually. This is due to the instrument's design for homogenous, light-transmitting liquid samples with a defined optical path. In contrast, solid lipstick surfaces scatter light and have surface irregularities, preventing consistent transmission. Additionally, a thin pH buffer layer above the sample allows for slight pigment molecule diffusion, resulting in the instrument primarily detecting colour changes in the buffer rather than the lipstick itself. This explains the strong linear correlation between absorbance peak and pH despite visible colour change on the lipstick being much weaker. The spectrophotometer, which measured reflected light rather than transmitted light, more accurately captured the lipstick's colorimetric behaviour, showing smaller but consistent absorbance peak wavelength shift, aligning with visual observations seen in swatches.

4.3 Biosensor Colorimetric Response in Lipstick Formulation

Although anthocyanin remained chemically active within the lipstick formulations, its colorimetric response was hindered significantly due to the hydrophobic beeswax-oil matrix. As a hydrophilic, water-soluble molecule, anthocyanin undergoes structural changes only in aqueous environments. In the lipstick, pigment particles are coated by hydrophobic melted wax and oil, forming a diffusion barrier that limits contact with the pH buffer, slowing activation and reduces the magnitude of colour change. Higher oil content further increases this hydrophobic barrier, reducing the number of pigment sites exposed to aqueous conditions and diminishing the colour transition [13]. Consequently, the colour shift observed on lipstick surfaces was considerably smaller than that in solution, though still measurable. All formulations exhibited an upward shift in absorbance peak after pH exposure, confirming that surface pigments remained responsive and capable of providing a biosensing signal. These findings indicate that the hydrophobic matrix limits colour expression but does not chemically degrade anthocyanin.

4.4 Overall Performance of Lipstick-Based Biosensor

The optimal formulation achieved balanced hardness, texture and spreadability comparable to commercial lipsticks while retaining a detectable colorimetric response within the salivary pH range. Although lipstick colour change was subtle compared to the biosensor in solution, it remained measurable visually and by spectrophotometry. The shift in absorbance peak representing colorimetric response was consistent with anthocyanin behaviour. While cosmetically effective, the modest colorimetric response may limit ease of interpretation for practical health monitoring.

5. Conclusion

This study successfully developed and evaluated a lipstick-based pH biosensor using red cabbage powder as a colorimetric indicator. By varying beeswax, mineral oil, and red cabbage powder (RCP) content, an optimal formulation of 1.5g beeswax, 1.0g mineral oil, and 1.0 g RCP was identified, achieving a balanced combination of hardness, smoothness and colour payoff suitable for cosmetic use. Mechanical testing confirmed a stable wax-oil network, while qualitative assessments showed good spreadability and uniformity. Both absorbance and reflectance measurements demonstrated that anthocyanin remained chemically active after incorporation, retaining biosensing functionality. Although the magnitude of colour shift was smaller than in aqueous solution, the lipstick produced a consistent upward shift in absorbance peak and a detectable darkening with increasing pH. Spectrophotometry proved more reliable than microplate absorbance for analysing solid samples. Overall, these findings confirm that a lipstick-based biosensor can function as a non-invasive and user-friendly platform for saliva pH sensing. While the colorimetric response is modest, the biosensing behaviour is reproducible and aligns with expected anthocyanin chemistry. Future work should focus on enhancing pigment accessibility to amplify the biosensor colorimetric response to improve ease of interpretation for real-world use.

6. Future Work

To improve viscoelasticity for better smoothness and structural strength at higher strains, further work could explore softer or blended waxes and additional components such as emollients, butters, or esters. Further optimisation of fabrication parameters, such as wax melting time, cooling duration and temperature, may also improve mechanical performance. Biosensor testing can be expanded using pH solutions prepared with artificial saliva containing physiological salts (e.g. KCl, NaCl, CaCl₂) to assess

potential interference effects on colorimetric response. Advanced optical methods such as diffuse reflectance spectroscopy with an integrating sphere [14] could be explored to obtain more accurate optical measurements of solid lipstick samples. Possible methods to enhance pigment accessibility could be explored, such as incorporating a small water-in-oil emulsion phase [15] to allow rapid penetration of aqueous fluid to pigment-rich regions. Finally, further performance metrics of the lipstick formulation should be investigated, such as reversibility, shelf life, and colour response time.

References

- [1] ScienceDirect (2025) *Biosensor - an overview | sciencedirect topics*. [Online]. 2025. ScienceDirect. Available from: <https://tinyurl.com/2u5hykru> [Accessed: 13 October 2025].
- [2] Bhalla, N., Jolly, P., Formisano, N. & Estrela, P. (2016) Introduction to biosensors. *Essays in Biochemistry*. [Online] 60 (1), 1–8. Available from: doi:10.1042/ebc20150001.
- [3] Mehrotra, P. (2016) Biosensors and their applications – A Review. *Journal of Oral Biology and Craniofacial Research*. [Online] 6 (2), 153–159. Available from: doi:10.1016/j.jobcr.2015.12.002.
- [4] Nakata, S., Arie, T., Akita, S. & Takei, K. (2017) Wearable, flexible, and multifunctional healthcare device with an ISFET chemical sensor for simultaneous sweat pH and skin temperature monitoring. *ACS Sensors*. [Online] 2 (3), 443–448. Available from: doi:10.1021/acssensors.7b00047.
- [5] Aydın, E.B., Aydın, M. & Sezgintürk, M.K. (2023) Biosensors for saliva biomarkers. *Advances in Clinical Chemistry*. [Online] 1131–41. Available from: doi:10.1016/bs.acc.2022.11.001.
- [6] Malon, R.S., Sadir, S., Balakrishnan, M. & Córcoles, E.P. (2014) Saliva-based biosensors: Noninvasive Monitoring Tool for Clinical Diagnostics. *BioMed Research International*. [Online] 20141–20. Available from: doi:10.1155/2014/962903.
- [7] Abedi-Firoozjah, R., Yousefi, S., Heydari, M., Seyedfatehi, F., et al. (2022) Application of red cabbage anthocyanins as pH-sensitive pigments in smart food packaging and sensors. *Polymers*. [Online] 14 (8), 1629. Available from: doi:10.3390/polym14081629.
- [8] Sun, S., Ku, Y.-H., Unsihuay, N., Florez, O., et al. (2024) BioCosMe: Lip-based cosmetics with colorimetric biosensors for salivary analysis using Deep Learning. *Proceedings of the 2024 ACM International Symposium on Wearable Computers*. [Online] 32–39. Available from: doi:10.1145/3675095.3676610.
- [9] Rigano, L. & Montoli, M. (2021) Strategy for the development of a new lipstick formula. *Cosmetics*. [Online] 8 (4), 105. Available from: doi:10.3390/cosmetics8040105.
- [10] Frothingham, S. (2018) *pH of Saliva*. [Online]. 27 September 2018. Healthline. Available from: <https://www.healthline.com/health/ph-of-saliva> [Accessed: 6 December 2025].
- [11] Frolova, Y., Sarkisyan, V., Sobolev, R., Makarenko, M., et al. (2022) The influence of edible oils' composition on the properties of beeswax-based Oleogels. *Gels*. [Online] 8 (1), 48. Available from: doi:10.3390/gels8010048.
- [12] Pan, S. & Germann, N. (2019) Thermal and mechanical properties of Industrial Benchmark Lipstick Prototypes. *Thermochimica Acta*. [Online] 679178332. Available from: doi:10.1016/j.tca.2019.178332.
- [13] Mattioli, R., Francioso, A., Mosca, L. & Silva, P. (2020) Anthocyanins: A comprehensive review of their chemical properties and health effects on cardiovascular and Neurodegenerative Diseases. *Molecules*. [Online] 25 (17), 3809. Available from: doi:10.3390/molecules25173809.
- [14] ThermoFisher (2023) *Solid-state materials: Analysis through UV-visible spectroscopic techniques*. [Online]. 2023. ThermoFisher. Available from: <https://tinyurl.com/455kczv2> [Accessed: 6 December 2025].
- [15] Louise (2025) *Water-in-oil emulsions: Mechanisms, applications, and Innovations*. [Online]. 7 April 2025. Elveflow. Available from: <https://tinyurl.com/2t5y59bh> [Accessed: 6 December 2025].

Automated Process Synthesis: Implementation and Evaluation of a Genetic Algorithm for Flowsheet Generation and Optimisation

Authors: Adjetej Samantha & Okoye Onyi

Department of Chemical Engineering, Imperial College London, U.K. December 2025

Abstract

This study evaluates the viability of a Genetic Algorithm (GA) for automated process flowsheet synthesis and optimisation. The proposed framework integrates Python with Aspen Plus via the Component Object Model (COM) interface, utilising the Non-dominated Sorting Genetic Algorithm II (NSGA-II) with hyperparameters derived from established literature. The methodology was validated through two distinct case studies: the Haber-Bosch process (Case 1) established algorithmic baseline performance, while the production of p-Xylene (Case 2) was employed to assess robustness under increased complexity. In Case 1, the algorithm successfully generated a Pareto front of optimal configurations, achieving an ammonia yield of 199 kmol/hr at a purity of 0.999. Case 2, which aimed to minimise deviation from target yields and maximise the purity of the ethylene by-product, successfully identified viable process topologies with a minimum deviation of 3.3015 kmol/hr and ethylene purity of 96.72%, although the global optimum was not definitively confirmed. To address computational bottlenecks, a surrogate model was implemented for candidate pre-screening; this mechanism effectively rejected unpromising simulations. The PSGA resulted in a deviation of 0.1599 kmol/hr and ethylene purity of 97.56%, as well as significantly reducing convergence time from 6.9 hours to 3.85 hours and lowering the required number of generations. These findings demonstrate that a GAs is a potent tool for discovering novel process architectures, though further development is required to ensure global optimality in complex systems.

Keywords: Genetic algorithm, Flowsheet generation, Aspen Plus & Optimisation.

1. Introduction & Background

The conventional approaches to conceptual process design heavily rely on domain knowledge and exhaustive simulation, leading to a restricted exploration of the vast design space. Design heuristics, such as the well-known hierarchical method for conceptual design by Douglas [1], as well as principles from Towler & Sinnott [2] and Seider et al. [3], provide systematic, human-driven methods for creating flowsheets that are applicable and inherently viable for industrial use. Given the time constraints of manual design, the full exploration of this design space, due to practicality, is not pursued. This limitation highlights the potential for automated machine learning techniques to discover novel and efficient flowsheets. Machine learning is poised to spearhead the automation of flowsheet generation, offering a solution that overcomes these bottlenecks. A genetic algorithm (GA) has the ability to navigate the complex design space, simultaneously handling both the structural selection of process units and the parametric optimisation of their operating conditions. This study aims to assess the viability of using this metaheuristic optimisation approach to rapidly generate and assess a multitude of feasible, high-performance flowsheets, leveraging the simulation environment of Aspen Plus to ensure the accuracy of each result. The findings of this work aim to lay the foundation for future studies, ultimately allowing for the integration of robust automated generation capabilities into industrial process design, significantly enhancing the work of professional engineers.

Aspen Plus is an essential tool in process engineering, aiding in both the development of process flow diagrams and high-fidelity process simulations [4]. With powerful thermodynamic models, and its ability to simulate a wide range of complex equipment units such as reactors, separators, and distillation columns; it has found applications extensively across chemical and petrochemical industries [5]. Aspen HYSYS and Aspen Plus are among the most referenced process simulators in academic literature, serving as the foundation for modelling, optimisation, control integration, and education [2][5]. The development of Aspen Hybrid will allow for the use of AI to aid in process design showing the industry shift towards automation [6]. While powerful on its own, the interconnectivity of Aspen Plus with external software like Excel, MATLAB, and Python (used in this study) has been shown to significantly improve research output and accuracy [7], allowing for remote operation of the Aspen simulation environment. However, the use of Aspen Plus in conceptual design is heavily reliant on prior experience and domain knowledge to ensure that a simulated process is feasible.

This manual design process is further complicated by the trade-offs between competing objectives. For example, maximising product purity and yield with separation optimisation have been proven to work against one another especially when closer to the global optimum [8]. These process-level goals interact with broader objectives like minimising economic cost or energy consumption. This frames conceptual design as a Multi-Objective

Optimisation (MOO) problem, characterised by a set of non-dominated solutions (a Pareto front) that represent optimal trade-offs. Hence suggesting the use of relevant optimisation principles within the scope of flow sheet generation. The scale of this problem is vast with the challenges involving both structural optimisation through the selection and sequencing of process and parametric optimisation through the manipulation of the wide range of operating conditions within each selected unit [9].

Recent machine learning applications in process design have shown promise but face distinct scalability and flexibility challenges. Mann et al. developed eSFILES, a hybrid AI framework for hierarchical flowsheet representation; however, its universal application is currently hindered by the intensive knowledge engineering required to build comprehensive process ontologies [10]. Conversely, approaches like Triantafyllou et al. (2024) utilise Bayesian Optimisation for rigorous operational tuning but are limited to fixed flowsheet structures and unfit for true structural synthesis [11]. While Reynoso-Donzelli et al. (2024) successfully applied Reinforcement Learning (PPO) for simultaneous generation and control, their reliance on neural network surrogates necessitates computationally expensive training, limiting solution accuracy to that of the pre-trained models [12]. Consequently, a critical gap remains for an algorithm capable of simultaneously optimising both flowsheet structure and operating conditions without reliance on rigid topologies or extensive pre-training.

When assessing the viability of such machine learning methods, a genetic algorithm (GA) is uniquely positioned as a low-computational and high-accuracy approach. By its nature, it can overcome the limitations inherent in these various methods for a hybrid synthesis approach. A GA is suited to this problem due to its "gene" and chromosome-based structure that can naturally encode both discrete and continuous variables simultaneously. This addresses the core challenge of this field by optimising the flowsheet unit selection and the continuous operating parameters concurrently within a single run. This provides a strong middle ground compared to other methods. Unlike the surrogate-based approaches, a GA does not require extensive, data-intensive model training. Instead, it works directly with the Aspen process simulator. The unique mechanism of a GA, which evolves an initial population of solutions, is highly effective. By manipulating genes representing both unit selections and their parameters, it explores the design space in parallel, making it robust and less likely to be trapped in a local optimum. When connected to Aspen, this framework simplifies the problem significantly. The GA treats the Aspen model as a black box fitness function, passing a set of genes into the simulation, and Aspen returns a "fitness" for each objective function.

The rigorous physical calculations and process constraints are all handled within the Aspen model itself, bypassing the need for complex mathematical reformulation.

2. Methodology

Genetic Algorithms (GAs) are meta-heuristic optimisation algorithms that computationally mimic the process of natural selection. By evolving a population of candidate solutions over multiple generations, GAs effectively explore complex search spaces to identify multiple optimal individuals. Each candidate solution within the population is represented by a unique set of chromosomes. These chromosomes function as the genotypic representation, where individual genes encode specific simulation parameters. When decoded, these parameters define the attributes of the system, i.e. the process flowsheet and determine its performance.

This study utilises the Non-dominated Sorting Genetic Algorithm II (NSGA-II), a widely adopted elitist multi-objective evolutionary algorithm [13]. NSGA-II classifies the population using a non-dominated sorting mechanism. Individuals are assigned a rank based on their dominance status; solutions that are not dominated by any other solution in the population form the first Pareto front (Rank 1), representing the set of optimal trade-off solutions. A critical feature of NSGA-II is its preservation of diversity through the calculation of "crowding distance." When comparing solutions of the same rank, the algorithm prioritises those located in less crowded regions of the solution space. This mechanism prevents genetic drift and ensures a uniform spread of solutions across the Pareto front. Consequently, NSGA-II was selected for this methodology due to its proven ability to generate diverse, high-quality alternative flowsheet configurations while reducing the risk of premature convergence to local optimum.

Aspen to Python Connection

The Aspen Plus model was connected to the python environment using the COM (Component Object Model) protocol. A custom Simulation class was utilised to manage the exchange between the two software platforms. This class interacts with the Aspen Plus hierarchical tree structure, using the pywin32 library to locate specific streams and blocks.

Key process variables were modified by navigating the simulation tree to the specific node and updating its value attribute. For example, optimising the feed temperature required the script to locate the specific stream element and overwrite the existing temperature value. While the underlying communication protocols were based on the methodology of Ricardez-Sandoval et al. [14] as well as Wolfgang [15], the interface was substantially expanded to

support the additional equipment and design heuristics proposed in this study.

The GA framework was designed to be adaptable to various chemical processes by adopting the hierarchical synthesis approach outlined by Douglas [1]. As depicted in Figure 2.1, the generated flowsheets follow a standard sequential superstructure: feed conditioning (temperature and pressure adjustment), reaction, and downstream separation. The algorithm manages these stages via a segmented genetic structure, where the separation synthesis is governed by a unique subset of chromosomes. This approach facilitates the simultaneous optimisation of reactor conditions and separation sequences within the same python script.

Validation of Methodology

Validation of the methodology was conducted via two case studies: a fundamental system used for algorithmic tuning (Case 1) and a complex multi-reaction system used to test performance limits (Case 2). In both instances, the simulation environment was pre-configured with the required physical properties and reaction stoichiometry to ensure thermodynamic accuracy. However, the simulation space was initialised as a blank slate containing no unit operations or stream connections. This constraint ensures that the Genetic Algorithm is responsible for generating the entire process flow diagram, validating its capability to generate viable flowsheets without prior structural guidance. Within the GA basic heuristics were encoded such as, penalties for invalid flow sheet units and non-converged solutions.

Case 1

Case 1 used the Haber-Bosch process to calibrate and validate the synthesis capabilities of the GA, with reaction kinetics and thermodynamic parameters sourced from literature [16] [17]. To isolate the algorithm's ability to generate correct unit decisions, the operating conditions (reactor temperature, pressure and feed ratio) were initially fixed at known optimal values hence the GA was tasked solely with selecting the appropriate unit operations for feed manipulation, reaction, and effluent separation. This case was critical for algorithmic development due to its physicochemical specificity. The high-pressure gas-phase

nature of the reactants dictates strict equipment requirements: a Plug Flow Reactor (PFR) is preferred over a CSTR to maximise conversion per pass; a Compressor is required rather than a pump due to the gaseous feed; and a Flash vessel is necessary for phase separation [18]. Therefore, the validity of the algorithm was defined by its ability to automatically converge on a flowsheet containing these three specific unit operations without user intervention. The algorithm was then used for the optimisation of the process parameters.

Case 2

Case 2 used the production of P-xylene as its process to be designed, with details of the process description found in the supplementary materials. This problem formulation introduced competing objectives and strict operational constraints, presenting a more challenging search landscape than the Haber-Bosch baseline. The initialisation protocol remained the same with Aspen Plus environment was pre-loaded with property methods and reaction stoichiometry. The algorithm executed a simultaneous synthesis and optimisation strategy, selecting unit operations while tuning their specific operating parameters. This approach allowed for the exploration of a wider variety of flowsheet configurations, testing the algorithm's ability to handle mixed-integer optimisation problems without getting trapped at a local optimum. Manual simulations and literature were used to evaluate the optimality of the produced solutions.

Within the optimisation in both cases operating parameters such as temperature and pressure were treated as discrete rather than continuous variables. This discretisation strategy was chosen to significantly reduce the search space and computational runtime, under the assumption that minor incremental fluctuations in these parameters would yield negligible improvements to the process.

Flowsheet Evaluation

To ensure the independence and reproducibility of simulation runs, a re-initialisation protocol was implemented. At the start of each evaluation cycle, the simulation environment was reset by purging all existing unit operations and material streams. The algorithm then constructed the specific flowsheet topology (units and streams) defined by the individual's chromosome. This

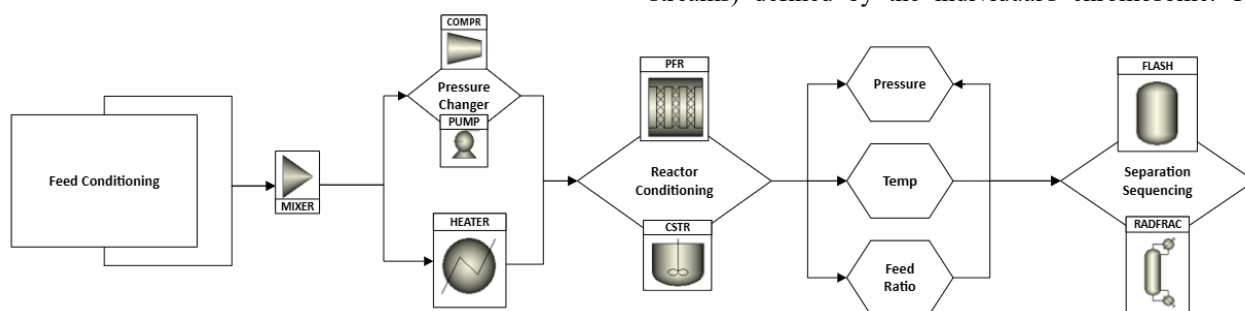


Figure 2.1: Basic flowsheet design

sequential process involved placing the selected pressure manipulation unit (pump or compressor), the reaction unit (CSTR or PFR), and the separation unit (Flash or Radfrac) at their specified conditions.

This was followed by the automated routing of material streams to establish the required connectivity between unit ports. Following the hierarchical approach suggested by Douglas [1] of initially dealing with the reaction system first, checkpoints were placed which invalidated flowsheets with defective reactor and pressure changer systems. These defective solutions were automatically assigned penalty values in the objective function to give them a low rank, making them less likely to be considered in subsequent generations. Parameter specification was directly encoded into the process units within the evaluate flowsheet definition. For separation units, a heuristic strategy was used to improve convergence stability.

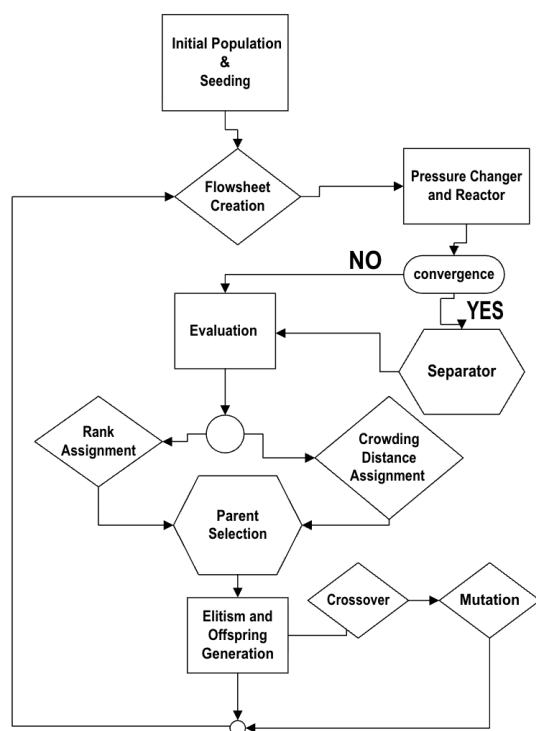


Figure 2.2: Genetic Algorithm operations flowchart

When initialising a rigorous distillation column (Radfrac), the system first ran a copy of the design through the shortcut distillation model (DSTWU) to estimate key design parameters. This included the theoretical number of stages as well as the distillate-to-feed ratio required for specified recovery targets. These outputs were then passed as input specifications to the Radfrac model, significantly reducing the likelihood of convergence failure in complex columns.

Following the construction of the individual, the Aspen Plus engine was invoked to solve the mass and energy balances. The method included a strict feasibility check: simulations returning convergence errors or physical

inconsistencies were immediately flagged. Though Aspen simulation warnings were not given the same severity of treatment as warning flowsheets can usually be manually fixed by editing the convergence requirements of the simulation. These non-converged solutions were assigned null fitness values for yield and purity, effectively acting as a penalty function that removed non-viable designs from the genetic pool. For successfully converged simulations, the interface extracted the relevant stream results to calculate the multi-objective performance metrics specific to each of the two cases.

Individual Creation

The initial population was generated by randomly creating individuals based on a predefined three-stage process. Each individual was encoded as a chromosome of three sequential genes: Pressure Changer, Reactor, and Separator. The Pressure Changer gene randomly selected a PUMP or a COMPRESSOR, with both initialised to a fixed target. The Reactor gene involved a choice between a PFR and CSTR. For this gene, continuous sizing parameters PFR diameter or CSTR volume were randomly initialised within their defined operating constraints. The final chromosomes governed the separation train, selecting between a simple two-phase Flash Separator (with a fixed operating temperature) and a Radfrac. If the Radfrac unit was selected, a heuristic-based Initialisation protocol was used, which automatically determined the light key and heavy key components based on component boiling points.

Assignment of Rank, Crowding distance & Selection

The evaluation phase of the genetic algorithm ends in the assignment of rank and crowding distance to each individual, which dictates their priority in selection for the next generation as indicated by the NSGA-II framework. Following the calculations of objective parameters, the individuals are sorted into a series of ranks. An individual is said to dominate another if it is superior in at least one objective and not inferior in any other objective and hence given the rank 1 (the pareto front). Once the first front is established, these individuals are temporarily excluded, and the domination analysis is repeated on the remaining pool to identify subsequent fronts to ensure the algorithm maintains a diverse set of solutions across the entire Pareto front, a crowding distance is calculated for every individual within each established front. Solutions with lower ranks are preferred, but when two individuals share the same rank the larger crowding distance is preferred for mating. Solutions on the edges of the front are automatically assigned an infinite distance to guarantee their preservation. By prioritising larger crowding distances, the algorithm prevents premature convergence to a single, localised region of the Pareto front, thereby preserving

diverse trade-off solutions between the conflicting objectives.

Mutation

The mutation phase introduced genetic diversity by randomly altering the genes of the offspring to prevent premature convergence. The mutation operator was applied to all genes using two distinct methods: Type Mutation and Parameter Mutation. Type Mutation, set at a 5% probability, involved the random change of the unit operation itself (e.g., PUMP to COMPRESSOR or PFR to CSTR). When this discrete change occurred, the associated parameters were immediately reset and re-initialised according to the rules of the new unit. For smaller, continuous adjustments, Parameter Mutation was applied with a 10% probability. This involved applying a constrained, random tweak to continuous variables (e.g., PFR diameter, CSTR volume). This adjustment was limited to $\pm 20\%$ of the parameter's total defined range to ensure the mutated value remained within physically realistic boundaries, thereby facilitating local fine-tuning of existing designs.

HYPERPARAMETERS	
TYPE MUTATION RATE	0.05
PARAMETER MUTATION RATE	0.1
CROSSOVER RATE	0.7
ELITISM RATE	0.1

Table 2.1: Hyperparameter decisions

Crossover

The generation of new flowsheets utilised the uniform cross over method, which was selected based on established literature as the preferred technique for maintaining diversity and mitigating linkage bias in process optimisation chromosomes [19]. Crossover between two parent flowsheets was executed with a probability equal to the crossover rate of 0.7. If triggered, the offspring's chromosome was constructed gene-by-gene. At each gene position, the gene made up of both the unit operation and its parameters was randomly inherited from either parent with a 50% probability, ensuring maximum mixing of the discrete choices and continuous parameters. If the crossover condition was not met parent 1 was directly copied to the next generation.

The overall structure of the Genetic Algorithm can be seen in figure 2.2, where the elitism rate denotes the percentage of the best individuals that are carried into the next generation.

The Genetic Algorithm parameters, including population size, crossover probability, and mutation rates, were selected based on standard evolutionary computing guidelines established by Goldberg [20], with adjustments made for the specific multi-objective nature of the problem.

3. Results and Discussion

Case 1

Problem Formulation

As previously discussed in section 2, the genetic algorithm is expected to favour flowsheets only comprising of a compressor, a plug flow reactor and a flash separator, which are consistent with the physical requirements of the Haber process. Furthermore, Le Chatelier's principle predicts that due to the equilibrium nature of the reaction, high temperature and high pressure would increase the yield of ammonia [21]. For a given reactant feed of 100 kmol/hr of nitrogen and 300 kmol/hr of hydrogen a perfect conversion would produce 200 kmol/hr of ammonia therefore maximising the yield as an Objective function (OF) will clearly demonstrate the effectiveness of the flowsheet. Pairing this OF with maximising the purity of the product stream ensures the optimal flowsheets produced align with industry practice as well as comprehensive view of how effective the flowsheet is. Finally, minimising the reactor volume is a conflicting OF which is required to understand the trade-offs that the pareto frontier characterises. To add, this OF also minimises CapEx which accounts – albeit vaguely - for the economical aspect of flowsheet design without requiring computationally expensive analysis. A more rigorous investigation using tools such as the Aspen Process Economic Analyser, would need to be conducted for a conclusive statement on the economic feasibility of flowsheets.

Unit decisions

Initially only unit optimisation was implemented, and data reveals that the frequency of incorrect unit (s) chosen by the GA drops from 43% in generation 1 to 9% in generation 3 and finally less than 3% in the final generation. Validating the algorithm can select the only feasible units, consistent with literature. The single anomaly observed in generation 3 occurred when the algorithm selected a Radfrac column in place of Flash unit. However, this case, and others, are quickly recognised and the corresponding flowsheet is assigned a low fitness ranking. Therefore, the Pareto front, as seen in figure 3.1.a, will remain stable and representative of the feasible design space in spite of slight deviations from the optimum chromosome, further highlighting the algorithms robustness. The pareto front obtained in figure 3.1.a is isolated for unit optimisation and the operating conditions were fixed at a reactor temperature of 450C, reactor pressure of 195 bar and a flash temperature of 100C. As expected, all the optimal flowsheets produced approximately the maximum yield (198.98 kmol/hr) and purity (0.999). The dominant trade-off was seen in reactor volume for negligible changes in yield and purity making the flowsheet with minimal reactor volume (0.0018 m³) the obvious optimal solution to reduce CapEx.

The trade-off between the reactor volume and maximising the yield and purity of ammonia is expected as increasing the volume will increase the residence time allowing the reaction to reach equilibrium.

Parameter Optimisation

In parameter optimisation, the dimensionality of design space increased exponentially as each gene was no longer binary, fundamentally altering the range of fitness within the population allowing for trade-offs to be conceptualised. To reduce computational time, the continuous variables were discretised and bound within predefined ranges.

The results of the Case 1 optimisation study are demonstrated in the Pareto Front in Figure 3.1.b.

When the operating conditions are included as decision variables, the range of the objective functions expand. Maximum purity produced from the optimal flowsheets, increased from 0.9999 to 1.0000, maximum yield increased from 198.9826 to 199.0111 kmol/hr, and minimum volume decreased from 0.0018 m³ to 0.0017 m³. Although the improvements are minimal, it is expected given the initial case had fixed operating parameters, which were the ideal conditions based on literature [16]. The contrast between the pareto fronts in figure 3.1.a-b visualises this finding. Figure 3.1a (discrete variables) shows a compressed front because of the discrete combinations at a single operating point. Figure 3.1b (continuous parameters) exhibits a pronounced curved and wider spread emphasising the algorithms ability to leverage the increased design space furthermore, the validity and effectiveness of the genetic algorithm. This is supported by the research done by

Pappas et al[22] by exploiting the extra degrees of freedom, and integrating design and operation, more competitive flowsheets are created and therefore a greater flexibility to balance different objective functions according to specific design and process specifications.

Convergence failures in Aspen Plus emerged during initial population generation when gene combinations violated the problem constraints. This issue was significantly amplified upon introducing continuous operating variables; the number of possible infeasible flowsheets increased from 4 to 600, highlighting the exponential growth in constraint violation risk as decision space dimensionality expanded. To improve the likelihood of locating the global optimum, the initial population size of 30 was increased to 50 furthermore, a larger number of generations was required. Although these adjustments were simple to implement, they significantly increased computational time and should be considered when extending the genetic algorithm to more complex flowsheet synthesis cases.

Case 2

Problem Formulation

The objective function was formulated to align with the design specifications detailed in the supplementary materials. The primary process goal was to achieve an annual production of 110,000 metric tonnes of p-xylene, corresponding to a molar flow rate of approximately 129 kmol/hr. While the sale of side products was unconstrained, the scope of this study was limited to the first two separation units as further separations required the use of extraction and crystallisation to deal with azeotropes that were not included in the units available in this study. Hence the heuristic strategy prioritised the removal of ethylene (the most volatile component) in the first column, followed by the recovery of xylenes from the column effluent in the second unit. Accordingly, the fitness of the proposed structure was defined by minimising deviation of the p-xylene yield relative to the 129 kmol/hr target ($Deviation = |129 - Y_{p-xylene}|$), and maximising purity of the ethylene product in the first separation column. To ensure feasibility and adherence to system constraints, penalty functions were applied to flowsheets exhibiting conversion rates outside the acceptable range. Parameters that were varied for optimisation included the reactor dimensions, pressure, temperature and the flash temperature. To address the complexities of the separation sequence, the genetic algorithm encoded the unit operations within the chromosome structure. Although the target products were pre-defined, the algorithm's gene representing the Radfrac block was permitted to vary the component 'cut point', and thereby dynamically determine the selection of the Light Key (LK) and Heavy Key (HK) components based on

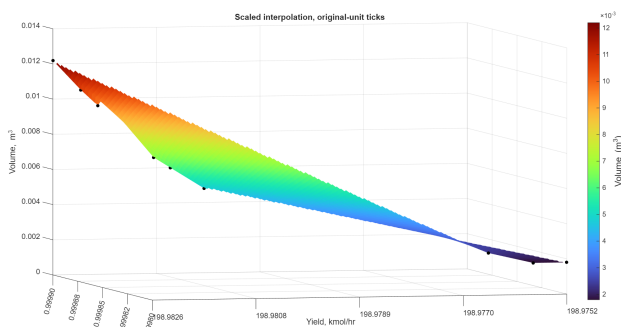


Figure 3.1 a) Pareto front for non-optimised Case 1

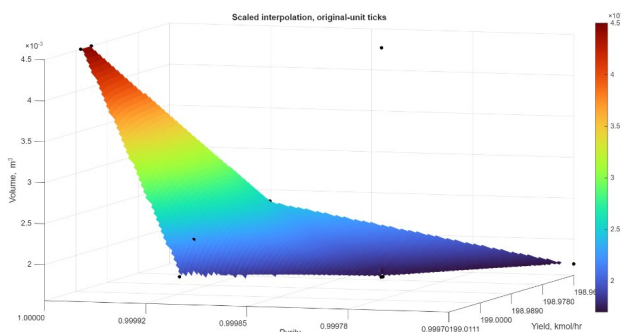


Figure 3.1 b) Pareto front for optimised Case 1

established heuristics [23]. Parameter mutations were increased to 0.4 due to the increase in the amount of parameters to be varied.

Results

Decision Analysis: The details of the Pareto front of flowsheets parameters as well as unit selection can be found in the supplementary material. The Genetic Algorithm (GA) successfully converged to a Pareto-optimal front of diverse process flowsheets after 50 generations. While standard heuristics for p-Xylene production, specifically those detailed by Liu et al. [24], advocate for a liquid-feed Pump, a Plug Flow Reactor (PFR) for continuous reaction, and a Flash-Distillation separation sequence, the GA explored a broader search space. Notably, the algorithm identified a dominant preference for pumps over compressors, with pumps selected in 60–95% of flowsheets across all generations as seen in figure 3.3.c However, in low-pressure differential scenarios (1 to 2 bar), the GA occasionally selected compressors. While thermodynamically feasible within the Aspen Plus simulation, such a configuration is mechanically infeasible for liquid feeds, resulting in immediate failure in real world operation. Aspen indicated this through "Liquid phase detected in compressor inlet" warnings [25] though within the code warnings were not penalised. This highlights a critical limitation in using standard thermodynamic simulations as black-box reward functions: without explicit mechanical constraints, the ML model optimises for the specified fitness functions without considering operational viability.

Reactor selection displayed a shift in evolutionary decisions. Early generations favoured PFRs, but the population stabilised to a close to even split between PFRs and CSTRs after generation 17 (figure 3.3.c). This suggests that while PFRs are preferred in engineering practice for their superior conversion-per-volume ratio, this heuristic was not encoded in the GA's objective function. Since the algorithm prioritised purity and yield not capital cost or equipment volume, it treated both reactor types as equally viable provided they met the output constraints. Regarding sizing, the GA successfully identified reactor dimensions capable of maximising output. However, manual validation runs revealed that a PFR with a length of 5.6 m (at 557°C and 6 bar) could achieve the required p-Xylene output with a smaller volume than the GA-derived solutions. This indicates that while the GA successfully maximised ethylene purity the Pareto front for reactor sizing could be further optimised by including volumetric minimisation as a specific objective as was done in Case 1.

The edges of the Pareto front revealed the distinct trade-offs. The highest purity consistent with Liu et al. [24] (96%) was achieved using a Flash followed by a RadFrac column

at high flash temperatures (30°C and 50°C). Conversely, the minimum deviation was observed at lower bounds: 2 bar pressure, 477°C, and a stoichiometric toluene-to-methanol ratio of 1.0. The use of uniform crossover, which randomly exchanges unit operations between parent

flowsheets, introduced high variance. While this prevented premature convergence by disrupting local minima, it also occasionally disrupted high-performing "building blocks" (e.g., compatible reactor-separator pairings). Furthermore,

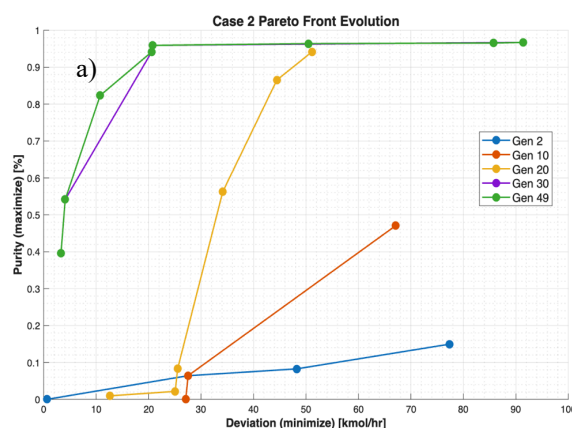


Figure 3.3 a) Pareto front Evolution for P-xylene production process

despite the inclusion of a flash heuristic, the GA occasionally generated suboptimal separation sequences. This behaviour is attributed to the objective function's focus on product purity rather than bottoms composition; as long as the p-Xylene product met target specifications, the algorithm did not penalise the energy-intensive presence of unreacted toluene or methanol in the bottoms stream.

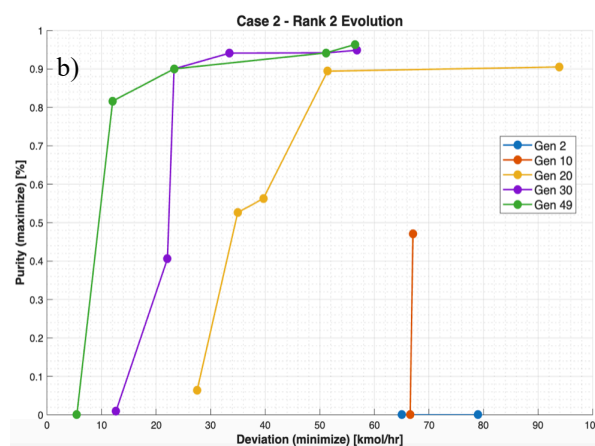


Figure 3.3 b) Rank 2 Evolution for P-xylene production process

Computational Performance and Convergence: The optimisation trajectory mirrors the Stochastic Superstructure Optimisation challenges described by Grossmann et al. [26], where complex parameter interactions often lead to local optima. Rather than evaluating parameters in isolation, the algorithm learned to

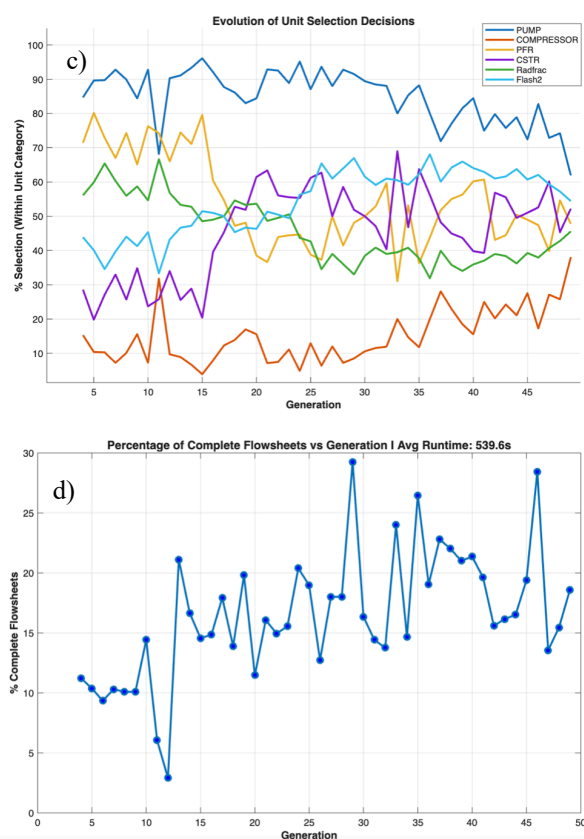


Figure 3.3 : c) Evolution for unit selection d) Evolution of Percentage of completed flowsheets

exploit multivariate interactions to maximise fitness. For instance, the system balanced higher flow rates with lower temperatures to maintain equilibrium, as evidenced by the solution achieving a purity of 0.8234 with a deviation of 10.8. Furthermore, incorporating economic and environmental constraints in future iterations would help filter the Pareto front, ensuring that selected solutions are not only theoretically optimal but industrially viable. The implementation of strict penalty functions for non-convergence or low yield significantly accelerated learning. As shown in Figure 3.3.a, the population achieved practical convergence within 50 generations, although rank-2 solutions continued to evolve (Figure 3.3.b). The percentage of fully converged flowsheets generally increased over time, though it fluctuated (e.g., dropping to 3% in Run 12) as depicted in figure 3.3.d. This fluctuation is inherent to the vast search space, where crossover operations frequently generate incompatible parameter sets. The average computational time was 536 seconds per generation. While the GA successfully generated viable starting points for process synthesis, the computational strain on the Aspen simulation engine limited runs to two generations per batch. To transition this approach from a theoretical tool to an industrial aid, future work must focus on improving the interface stability between the Python-based GA and the chemical simulator, potentially by integrating cost and environmental constraints to filter the Pareto front for industrial applicability. Though, the algorithm is more than sufficient in providing a range of

options with parameters that can be further optimised; acting as a good base for the initial steps of flowsheet design for example within distillation columns parameters such as reflux ratio could be varied to increase purity.

Prenatal Screening- Genetic Algorithm (PS – GA)

Prenatal Genetic Screening (PGS) mechanism was integrated into the Genetic algorithm [27]. This hybrid approach used a surrogate model to estimate the viability of potential flowsheet without the computational expense of a full Aspen Plus simulation. The surrogate system comprised of two distinct Multi-Layer Perceptron (MLP) regressors (implemented via Scikit-learn), specifically trained to predict the 2 objectives of the ethylene purity and xylene yield deviation. The PGS-GA functions through a three-stage iterative process. Chromosome Encoding transforms flowsheet configurations including discrete unit choices (e.g., PFR vs. CSTR) and operating parameters (e.g., temperature) are flattened into numerical feature vectors through Heterogeneous Data Serialisation. Before simulation, generated offspring are evaluated by the surrogate model. A "prenatal" filter is applied; only candidates with predicted performance metrics satisfying specific thresholds [40% ethylene purity and 60 kmol/hr deviation] are selected for rigorous evaluation within the Aspen simulation environment. Candidates failing this screening are discarded immediately, preventing wasted computation on non-viable designs. Data obtained from successful rigorous simulations are then dynamically appended to the training set. The MLP regressors are iteratively retrained at every generation, allowing the surrogate model's predictive accuracy to evolve simultaneously with the population. This methodology effectively reduces the search space by filtering out low-quality or non-converging solutions early, significantly optimising the total runtime required for flowsheet generation.

Results

Full details on the final pareto front of the PSGA can be found in the supplementary materials. The implementation of the PGS-GA successfully reduced the computational time required for Rank 1 convergence from approximately 6.9 hours to 3.85 hours, taking only 28 generations. The initial convergence rate was significantly higher, with the majority of the Pareto front identified by generation 10 (Figure 3.4.a and 3.4.b) and Rank 2 converging shortly after. Although Shaffie et al [27] saw their PGS-GA required more generations than a standard Genetic Algorithm (GA), this discrepancy is likely due to their use of a single hidden-layer neural network. In contrast, this study employed a double hidden-layer architecture (15 and 8 neurons) for both purity and deviation models. This architecture provided sufficient complexity to mimic the process dynamics without hindering runtime.

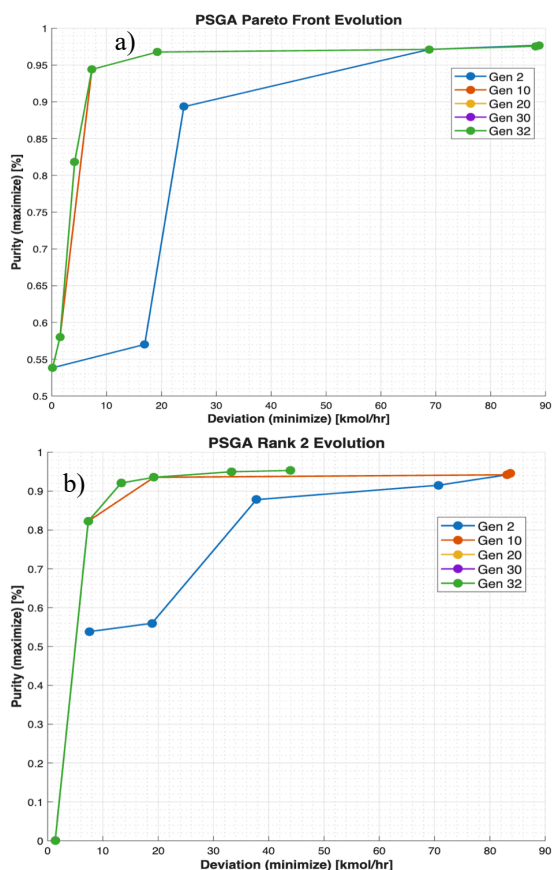


Figure 3.4 a) PSGA Pareto front Evolution for P-xylene production process b) PSGA Rank 2 Evolution for P-xylene production process

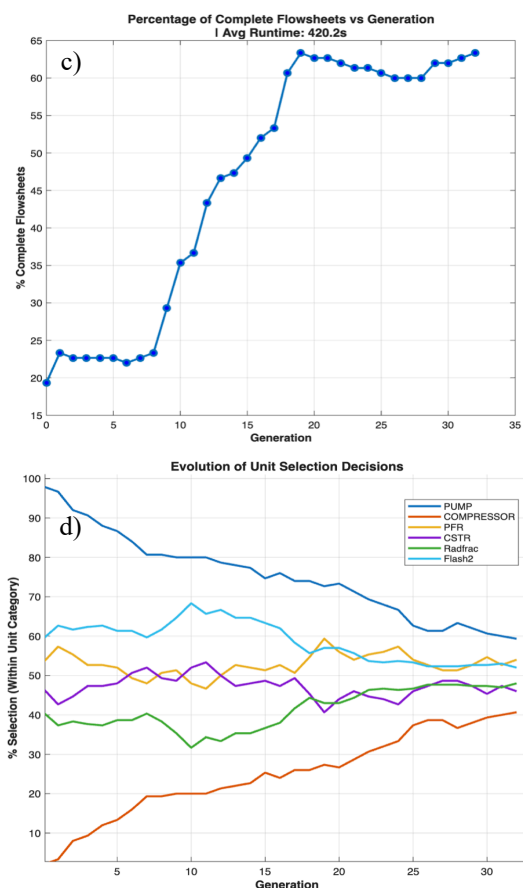


Figure 3.4 c) PSGA Evolution of Percentage of completed flowsheets d) PSGA Evolution for unit selection

While the PGS-GA yielded fewer non-dominated solutions, the quality of candidates improved. The maximum ethylene purity reached 97.56% (compared to the GA maximum of 96.72%). Furthermore, deviations from the target p-xylene production were consistently lower, with the best individual achieving a deviation of 0.1599 kmol/hr, though at a lower ethylene purity of 54%. The surrogate model screening proved highly effective in filtering the design space for viable flowsheets, by utilising a hybrid training set of 918 complete solutions over all 32 generations. This learning capability is evident in Figure 3.4.c; the simulation completion rate rose from 20% to 65% within 17 generations, double the rate of the standard GA. Of these completed flowsheets, 100% met the minimum 40% ethylene purity constraint, though only 48% satisfied the maximum deviation criteria (<60) by generation 32. Regarding decision variables, the surrogate model facilitated a more deterministic selection process (Figure 3.4.d), reducing the sporadic decision making due to stochastic search observed in the standard GA. Reactor and separator selection ratios remained stable throughout the evolution. Pressure changers converged toward an even split between pumps and compressors. Notably, five of the eight Rank 1 individuals utilised compressors despite pump systems generally yielding higher success rates. This suggests that the early convergence of the Pareto front, combined with ANN screening, may have biased the population by rejecting pump configurations deemed unlikely to succeed based on early training data. Ultimately, the PGS-GA aided the algorithm in producing marginally superior candidates. While it faced similar industrial feasibility challenges to the standard GA, the PGS-GA is arguably more applicable due to the significant reduction in runtime while producing flowsheets suitable for further optimisation.

4. Conclusion

This report presents the robust capability of the genetic algorithm (GA) for synthesising optimal process flowsheets through the integration of Aspen Plus with Python via the COM interface. Case 1 established the algorithm's stability in solving probabilistic models, leveraging unit operation decisions as discrete choices to converge on configurations consistent with engineering heuristics (Stops et al., 2022) [28]. When operating conditions were subsequently optimised, the Pareto front shifted towards improved performance; although the gains in objective functions were incremental, they demonstrate that the algorithm successfully exploits additional degrees of freedom to offer greater flexibility in balancing competing objectives.

Case 2 addressed the computational bottleneck by using a hybrid Proxy-Surrogate GA (PS-GA). By screening flowsheets before expensive simulation, this approach

markedly reduced computational time and the number of generations required, a notable insight for scaling to larger problems. Aligning with Shafiee et al. (2017)[27], the optimal solution was of superior quality. However, while the PSGA is more practically applicable for industry, systematic unit selection patterns were noted, suggesting that training data may introduce selection bias.

Across all cases, the algorithm reliably converged towards sensible configurations, verifying its use as a screening tool. However, thermodynamic and mass-balance constraints resulted in convergence failures within Aspen Plus when gene combinations violated physical limits. While measures were successfully added to improve global optimisation, they increased computational time, highlighting the need to embed mechanical constraints directly into the fitness function.

To advance this methodology from a theoretical tool to industrial practicality, future work should integrate cost and environmental constraints. Furthermore, while the present work employed a maximum of two separators, future investigations should allow the algorithm to optimise the separation train itself, including the number, type, and sequencing of units. This would enable the discovery of economically superior strategies, such as reactive distillation or pressure-swing absorption hybrids (Meng et al., 2024)[29] that are not immediately apparent from conventional engineering heuristics (Douglas, 1985)[1].

Acknowledgements

Thank you, God, Jesus and the Holy Spirit. Alice & Emmanuel Adjetej and Ejike & Uzo Okoye our lovely parents, for funding this degree! Thank you to the all the gals for giving us happiness outside our degree. Love you all, MWAH xxx

5. References

- [1] Douglas, J.M. (1985). A hierarchical decision procedure for process synthesis. *AIChE Journal*, 31(3), pp.353–362. doi:https://doi.org/10.1002/aic.690310302.
- [2] Seider, W.D., Lewin, D.R., Seader, J.D., Soemantri Widagdo, (Chemical Engineer, R Gani and Ka Ming Ng (2017). *Product and process design principles : synthesis, analysis, and evaluation*. New York: John Wiley & Sons Inc.
- [3] Towler, G. and Sinnott, R. (2008). *Chemical engineering design : principles, practice and economics of plant and process design*. [online] Amsterdam ; Boston: Elsevier/Butterworth-Heinemann. Available at: https://core.ac.uk/download/pdf/143491361.pdf.
- [4] Juma Haydari (2019). *Chemical process design and simulation : Aspen Plus and Aspen HYSYS applications*. New York, N.Y.: Aiche, The Global Home Of Chemical Engineers ; Hoboken, Nj, Usa.
- [5] Jana, A.K. (2014). *Process simulation and control using aspen*. Delhi: Phi Learning.
- [6] www.aspentech.com. (n.d.). *Hybrid Artificial Intelligence Models | AspenTech*. [online] Available at: https://www.aspentech.com/en/solutions/aspen-hybrid-models.
- [7] Bartolome, P.S. and Gerven, T.V. (2022). A comparative study on Aspen Hysys interconnection methodologies. *Computers & Chemical Engineering*, p.107785. doi:https://doi.org/10.1016/j.compchemeng.2022.107785.
- [8] Gernaey, K.V., Huusom, J.K. and Gani, R. (2015). *12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering*.
- [9] Boukouvala, F., Misener, R. and Floudas, C.A. (2016). Global optimization advances in Mixed-Integer Nonlinear Programming, MINLP, and Constrained Derivative-Free Optimization, CDFO. *European Journal of Operational Research*, [online] 252(3), pp.701–727. doi:https://doi.org/10.1016/j.ejor.2015.12.018.
- [10] Mann, V., Sales-Cruz, M., Gani, R. and Venkat Venkatasubramanian (2024). eSFILES: Intelligent process flowsheet synthesis using process knowledge, symbolic AI, and machine learning. *Computers & chemical engineering*, 181, pp.108505–108505. doi:https://doi.org/10.1016/j.compchemeng.2023.108505.
- [11] Triantafyllou, N., Lyons, B., Bernardi, A., Chachuat, B., Kontoravdi, C. and Papathanasiou, M.M. (2024). Comparative assessment of simulation-based and surrogate-based approaches to flowsheet optimization using dimensionality reduction. *Computers & Chemical Engineering*, [online] 189, p.108807. doi:https://doi.org/10.1016/j.compchemeng.2024.108807.
- [12] Reynoso-Donzelli, S. and Ricardez-Sandoval, L.A. (2024). An integrated reinforcement learning framework for simultaneous generation, design, and control of chemical process flowsheets. *Computers & Chemical Engineering*, [online] 194, p.108988. doi:https://doi.org/10.1016/j.compchemeng.2024.108988.
- [13]]Kumar, M. and Guria, C. (2017). The elitist non-dominated sorting genetic algorithm with inheritance (i-NSGA-II) and its jumping gene adaptations for multi-objective optimization. *Information Sciences*, 382-383, pp.15–37. doi:https://doi.org/10.1016/j.ins.2016.12.003.
- [14] Group, R. (2025). *Luis Ricardez-Sandoval Research Group / Flowsheet desing*. GitLab. [online] GitLab. Available at: https://git.uwaterloo.ca/ricardez_group/flowsheet-desing
- [15] Ten Hagen, R.W. (2022). *AspenPlus Python Interface*. [online] GitHub. Available at: https://github.com/YouMayCallMeJesus/AspenPlus-Python-Interface.
- [16] Modak, J.M. (2002). Haber process for ammonia synthesis. *Resonance*, 7(9), pp.69–77. doi:https://doi.org/10.1007/bf02836187.
- [17] Md. Burhan Kabir Suhan, Md. Naimur Rahman Hemal, M.A.A. Shoukat Choudhury, Md. Ali Akkas Mazumder, Mohaiminul Islam, Optimal Design of Ammonia Synthesis Reactor for a Process Industry, Journal of King Saud University - Engineering Sciences, Volume 34, Issue 1, 2022, Pages 23-30, ISSN 1018-3639, https://doi.org/10.1016/j.jksues.2020.08.004.
- [18] Rosbo, J.W., Jensen, A.D., Jørgensen, J.B., Sigurd Skogestad and Huusom, J.K. (2025). Optimisation of a Haber-Bosch Synthesis Loop for PtA. *Systems and Control Transactions*, [online] 4, pp.1126–1132. doi:https://doi.org/10.69997/sct.122254.
- [19] Z Kravanja and Miloš Bogataj (2016). *26th European Symposium on Computer Aided Process Engineering*. Amsterdam, Netherlands: Elsevier.
- [20] Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional.
- [21] Bashyal, J. (2023). *Haber's Process: History, Advantages, Disadvantages Detailed Process With The Ammonia Synthesis*. [online] scienceinfo.com. Available at: https://scienceinfo.com/habers-process-history-synthesis-ammonia/.
- [22] Pappas, I., Kenefake, D., Burnak, B., Avraamidou, S., Ganesh, H.S., Katz, J., Diangelakis, N.A. and Pistikopoulos, E.N. (2021). Multiparametric Programming in Process Systems Engineering: Recent Developments and Path Forward. *Frontiers in Chemical Engineering*, 2. doi:https://doi.org/10.3389/fceng.2020.620168.
- [23] Henley, E.J., Seader, J.D. and Roper, D.K. (2011). *Separation process principles*. 3rd ed. Hoboken, N.J. Wiley.
- [24] Liu, J., Yang, Y., Wei, S., Shen, W., Rakovitis, N. and Li, J. (2018). Intensified p-Xylene Production Process through Toluene and Methanol Alkylation. *Industrial & Engineering Chemistry Research*, 57(38), pp.12829–12841. doi:https://doi.org/10.1021/acs.iecr.8b00681.
- [25] Inc, A.T. (2000). *Aspen Plus ® Aspen Plus User Guide*. [online] Available at: https://web.ist.utl.pt/ist11038/acad/Aspen/AspUserGuide10.pdf.
- [26] Grossmann, I.E. and Guillén-Gosálbez, G. (2010). Scope for the application of mathematical programming techniques in the synthesis and planning of sustainable processes. *Computers & Chemical Engineering*, [online] 34(9), pp.1365–1376. doi:https://doi.org/10.1016/j.compchemeng.2009.11.012.
- [27] Shafiee, A., Nomvar, M., Liu, Z. and Abbas, A. (2017b). A new genetic algorithm based on prenatal genetic screening (PGS-GA) and its application in an automated process flowsheet synthesis problem for a membrane based carbon capture case-study. *Chemical Engineering Research and Design*, [online] 128, pp.265–289. doi:https://doi.org/10.1016/j.cherd.2017.10.009.
- [28] Stops, L., Leenhouts, R., Gao, Q. and Schweidtmann, A.M. (2022). Flowsheet generation through hierarchical reinforcement learning and graph neural networks. *AIChE Journal*, [online] 69(1). doi:https://doi.org/10.1002/aic.17938.
- [29] Meng, J., Wang, C., Bai, Y., Dong, Y. and Du, J. (2024). Automatic Design of Extractive Distillation Sequence for a Multicomponent Azeotropic System. *Industrial & Engineering Chemistry Research*, 63(5), pp.2294–2306. doi:https://doi.org/10.1021/acs.iecr.3c03575

Evaluation of the Natural Ortho-Para Hydrogen Conversion Kinetics

Jaace Chern, Wan Iqbal Wan Rawi

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Accurate modelling of ortho-para (OP) hydrogen conversion and its reaction kinetics is essential for the design of efficient liquefaction and storage processes. In this paper, natural OP conversion kinetics were investigated between 200 and 300 K at pressures up to 20 MPa using two complementary techniques: speed of sound measurement in a cylindrical acoustic resonator and rotational Raman spectroscopy in a stainless steel cell equipped with a sapphire window. The speed of sound method indirectly infers the OP composition from resonance frequencies, whereas Raman spectroscopy provides a direct measure of rotational level populations. A second-order homogeneous autocatalytic rate model (Milenko model) was fitted to both datasets to obtain combined rate constants. At 20 MPa, Raman-derived values at 298 K acquired were 0.0275 hr^{-1} and 0.0252 hr^{-1} , which agreed closely with the extrapolated model, which is 0.0257 hr^{-1} . However, k'' values obtained from speed of sound data — 0.0157 hr^{-1} and 0.0182 hr^{-1} — exhibited larger deviations from the model's prediction of 0.0263 hr^{-1} . Although the Raman data are consistent with Milenko's homogeneous conversion mechanism at elevated temperatures, the results do not provide a definitive validation of the model above 120 K. Hence, the extrapolated predictions above 120 K should be treated only as qualitative guidance, rather than evidence that the model remains valid at these temperatures.

Keywords: Catalysis, Energy, Thermodynamics Process, Ortho-Para conversion, Kinetics, Hydrogen Liquefaction

1 Introduction

1.1 Background

The constant growth in global population, and its associated industrial activities, has contributed to detrimental effects, especially on greenhouse gases (GHGs) and carbon emissions.¹ Hence, other energy vectors need to be considered to reduce the public dependency on traditional energy sources such as fossil fuels and shift the focus towards renewables and carbon-free energy carriers.² Hydrogen, H_2 , has been emerging as one of the most promising solutions available, even though it has come along with multiple discrepancies; hence, the large-scale implementation and use of liquefied H_2 is yet to be possible. Large-scale hydrogen liquefaction is hindered by economic constraints, cryogenic boil-off losses, safety concerns associated with high-pressure hydrogen operations and inherent flammability, and process scale-up constraints.³

To enable successful liquefaction and minimal regasification of liquefied H_2 , accurate monitoring and kinetic

model development of the conversion between OP H_2 is paramount. Normal H_2 is referred as H_2 at ambient conditions, which is a mixture of 75% ortho- H_2 and 25% para- H_2 . The equilibrium composition of the two nuclear spin isomers is a temperature-dependent property,⁴ hence close monitoring of the operating temperature and pressure is essential.

Hydrogen liquefaction is essential for practical storage and long-distance transportation. Liquid H_2 typically requires refrigeration to 20 K, which leads to more than 99.8% of para- H_2 . However, the corresponding OP conversion reaction is highly exothermic due to the significantly different internal energies.² This subsequently leads to liquid boil off, and the energy penalty incurred onto this reaction is significantly high.⁵ Furthermore, as the transition from the ortho-to-para state is quantum mechanically forbidden in an isolated molecule, the natural OP conversion takes days or weeks to achieve completion.² Hence, a catalyst-like material must be introduced to enhance cooling and liquefaction kinetics.²

In our study, a ferrous oxide-based catalyst, Ionex®

Type OP catalyst, is used to increase the rate of OP conversion in a practical time frame. The corresponding rate expressions are derived for different operating conditions under more achievable temperatures within the range of 200 K and 300 K. The OP ratio is determined by measuring the speed of sound and analysing Raman spectrum, and the results are cross-validated to ensure their reliability.

1.2 Objectives

The objectives for this study are summarised as follows:

- Calculate OP spin isomer composition at equilibrium as a function of temperature.
- Calculate the speed of sound as a function of OP composition by using REFPROP to verify that the speed of sound is a linear function of ortho-H₂ mole fraction at constant temperature and pressure.
- Measure natural conversion rate at various constant temperature-pressure conditions by analysing the speed of sound of H₂ and Raman spectrum, and cross validate the results obtained from both approaches.
- Derive the time constants for each investigated case and extend the kinetic model to room temperature.

2 Theory

2.1 Physical Properties of H₂

Multiple physical properties of H₂ are identified as temperature-dependent properties, such as thermal conductivity and speed of sound, and are well established as indirect measurement methods of the isomer composition.² The OP ratio is further inferred from these temperature-dependent physical properties and requires calibration of the measurement system.² The configurations of the ortho and para spin isomers are shown in Figure 1. For this study, the speed of sound is analysed by using a cylindrical acoustic resonator, which is discussed further in Section 3.3.

Some of the thermal properties of H₂, including triple point, normal boiling point, and critical point are summarised in Table 1.⁶

Table 1: Summary of Thermophysical Properties of H₂

Property	Temperature / K	Pressure
Triple Point	13.8	7 kPa
Normal Boiling Point	20.3	0.1 MPa
Critical Point	32.9	1.3 MPa

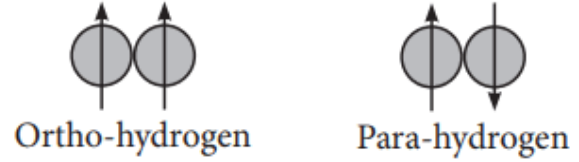


Figure 1: Configuration of Ortho and Para Hydrogen Spin Isomers

At a cryogenic temperature of 20 K, approximately 99.9% of para spin isomer is present. At 300 K, the composition of OP isomer is 75% and 25%, respectively.⁷

2.2 Temperature Definition

As mentioned in Section 2.1, the OP composition is highly dependent on temperature influences, hence an accurate definition of temperature is essential. The temperature referred in the sound speed measurement is the temperature of the acoustic resonator. The temperature of the catalyst bed is only altered to shift the equilibrium composition of normal H₂, hence accurate measurement of temperature at the catalyst bed is not required. On the other hand, the temperature of the Raman spectrum measurements is measured by using a temperature sensor located near the Raman measurement setup, which measures approximately the room temperature, and it was assumed to be the same as the Raman cell temperature. The temperature of the acoustic resonator and Raman measurement setup were then used for the deduction of rate constants.

As previous literature only investigated the temperature ranging approximately 100 K, it is essential to investigate the conversion kinetics at less cryogenic conditions, which ranges from 200 K to 300 K. The reported data are often old, and the uncertainties are not specified.⁶ Hence, the determination of rate equations for a more attainable temperature range with a more reliable approach is essential for large-scale implementation.

2.3 Modified Boltzmann Distribution

The original Boltzmann distribution is defined as shown in Equation 1. N_i is the number of molecules in state i , N is the total number of molecules available, g is the degeneracy of molecule, E is the molecule's energy, k is the Boltzmann's constant and T is the thermodynamic temperature. It is important to note that

the Boltzmann distribution provides the relative population of states instead of the energy levels.⁸

$$\frac{N_i}{N} = \frac{g_i \exp\left[-\frac{E_i}{kT}\right]}{\sum_j g_j \exp\left[-\frac{E_j}{kT}\right]} \quad (1)$$

By considering a rigid linear rotor model, which is applicable for H₂, E_J and g_J can be redefined as shown in Equation 2 and 3. Hence, the Boltzmann distribution equation, which only considers rotational contribution, can be rewritten as shown in Equation 4. θ_R is the rotational characteristic temperature of H₂ (which is 87.6 K) and J is the rotational energy level.⁸

$$E_J = J(J+1) \theta_R k \quad (2)$$

$$g_J = 2J + 1 \quad (3)$$

$$\frac{N_J}{N} = \frac{(2J+1) \exp\left[-J(J+1) \frac{\theta_R}{T}\right]}{\sum_{J=0}^{\infty} (2J+1) \exp\left[-J(J+1) \frac{\theta_R}{T}\right]} \quad (4)$$

H₂ molecules can only attain certain J levels depending on their isomer configuration. Para-H₂ can only attain even J states and ortho-H₂ can only attain odd J states.³ An extra degeneracy factor which contributed by the nuclear spin degeneracy, shall also be considered, as para-H₂ exists as a singlet nuclear spin molecule while ortho-H₂ exists as a triplet nuclear spin molecule.³

Hence, the OP spin isomer ratio could be estimated by using the modified Boltzmann distribution as shown in Equation 5, where x_p is the para-H₂ mole fraction and x_o is the ortho-H₂ mole fraction.⁶

$$\frac{x_p}{x_o} = \frac{1}{3} \frac{\sum_{J=0,2,4,\dots} (2J+1) \exp\left[-J(J+1) \frac{\theta_R}{T}\right]}{\sum_{J=1,3,5,\dots} (2J+1) \exp\left[-J(J+1) \frac{\theta_R}{T}\right]} \quad (5)$$

Based on the Taylor series derived from the Boltzmann distribution, higher orders of J correspond to more accurate results. Based on the analysis, when $\Delta J = 6$, the curve reaches a plateau, indicating that any ΔJ above 6 contributes minimally to the Boltzmann distribution. Hence, it is unnecessary to consider scenarios with $\Delta J > 6$. Figure 2 shows the $\frac{x_p}{x_o}$ calculated at different temperatures by using Equation 5 and provides first estimations for further experimental analysis.⁷

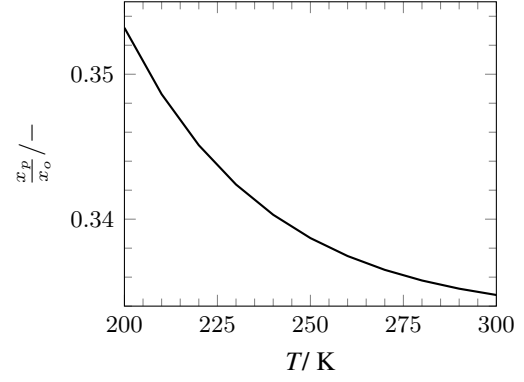


Figure 2: $\frac{x_p}{x_o}$ at Different Temperatures

2.4 Speed of Sound

The speed of sound of H₂ is calculated by using the acoustic model, which utilises the resonance frequency, f_{200} , of the second longitudinal mode of oscillation.⁹ As all the odd-numbered longitudinal modes are coupled to the cylindrical resonator's vibration, and low-frequency radial modes are eliminated as an effect of the resonator design, these requirements strictly limit the operation to only the second longitudinal mode.¹⁰

The speed of sound calculation follows Equation 6 under an idealised resonator design with rigid walls and negligible viscous and thermal boundary layer effects. c represents the speed of sound in the gas, f_{200} is the resonance frequency of the second longitudinal mode of oscillation, and L is the cavity length, which is 60 cm for the resonator design in our study.⁹

$$c = f_{200} \times L \quad (6)$$

However, under a real-life setting, corrections are applied to accommodate the effects of the viscous and thermal boundary layers in the gas at the cavity wall, as represented by Δf_{v-t} , and the presence of two identical end plates, as represented by Δf_{end} . The Equation 6 is modified and given by Equation 7 and Equation 8. Δf_{end} arises from the non-rigid end plates which served as electro-acoustic transducers and were used to both generate and detect the resonance frequency. This effect is accounted for by scaling the f_{end} term by a factor of two.⁹

$$c = [f_{200} - \Delta f] \times L \quad (7)$$

$$\Delta f = 2 \Delta f_{end} + \Delta f_{v-t} \quad (8)$$

2.5 Raman Spectroscopy

In Raman spectroscopy, the sample is exposed to photons of same energy level from the monochromatic radiation source. It is used to study molecular vibrations and rotations, and the molecule must be anisotropically polarisable.⁸ By analysing the frequency or the Raman shift of the peak occurring, it is possible to determine the vibrational and rotational motions that contribute to the peak.

When the incident photons are scattered onto the sample molecule, the molecule can undergo three types of scattering, namely Rayleigh, Stokes and Anti-Stokes scattering. Rayleigh scattering occurs when the incident photons are scattered elastically, which happens to most of them. 1 in 10^7 photons are scattered inelastically and either undergo Stokes scattering or Anti-Stokes scattering. In Stokes scattering, the photons lose energy to the incident molecule, and the molecule becomes excited. The photons gain energy from the incident molecule by Anti-Stokes scattering, and the emerging radiation has higher frequency.⁸ The energy level representation of different types of scattering is shown in Figure 3.

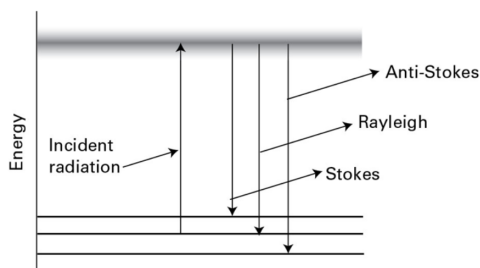


Figure 3: Energy Level Changes in Rayleigh, Stokes and Anti-Stokes Scattering⁸

As H_2 is classified as homogeneous diatomic linear molecule, it consists of both vibrational and rotational motions, and hence the contributions from both motions need to be considered separately in Raman analysis. However, for temperatures below 1000 K, the H_2 molecule is assumed to be always in ground vibrational state ($v = 0$). Hence, vibrational Raman spectroscopy is not considered.² All linear molecule, including H_2 , are rotationally Raman active, and this is exploited as most linear molecules are inaccessible to microwave spectroscopy.⁸ The population of each J state follows a Boltzmann distribution as discussed in Section 2.3.

2.5.1 Rotational Raman Spectra

For a linear rotor such as the H_2 molecule, the specific rotational Raman selection rules are established as shown in Equation 9, which are established based on the transitional dipole moment. $\Delta J = 0$ corresponds to Rayleigh scattering as it does not lead to a shift in frequency of the scattered photons. The Raman analysis in this study focuses solely on Stokes scattering, as Rayleigh scattering produces no Raman shift; therefore, no Raman peaks could be observed.

$$\Delta J = 0, \pm 2 \quad (9)$$

3 Methodology

3.1 Materials

Table 2: Description of Materials Used in the Experiments

Materials	Type	Purity / %	Supplier
Hydrous Ferric Oxide	Ionex® Type OP catalyst	100	Molecular Product Inc
Hydrogen	Gas	99.9995	BOC

3.2 Catalyst Activation

Activation of the catalyst is necessary before use to eliminate any adsorbed moisture and volatile contaminants that could otherwise poison the active sites. Due to the absence of on-site heating equipment for in-situ activation, the catalyst was pretreated in a vacuum oven at 160°C while establishing a dry nitrogen flow at the same time. The sample was evenly distributed on a glass dish and left heated overnight in accordance with Zhou et al.,¹¹ which recommends activation at 160°C for 8 to 12 hours. After the thermal treatment, the catalyst was allowed to cool down for approximately 3 hours before being immediately transferred to the reactor and sealed to minimise atmospheric exposure. The reactor was purged with dry H_2 gas once reattached to the experimental setup.

3.3 Apparatus

The setup consisted of a bench-mounted assembly that incorporated a series of valves (V1-V6), a hand pump (SITEC Hand), a control valve (CV1), pressure indicators (P1-P3), and the reactor (catalyst bed). P1 measures the gauge pressure while P2 and P3 measure the absolute pressure. All pressures measured by the indicators were reported in bar. The inflow of H_2 gas was controlled by a manual valve (GSV), and the

venting to the fume cupboard was controlled by V5 or V6 and the vacuum pump.

For speed of sound measurement, a cylindrical acoustic resonator was used based on the design of Ruffine and Trusler⁹ as shown in Figure 4, which was immersed in an ethanol or ethylene glycol-water bath. For safety precautions, the ethanol bath was layered with nitrogen gas due to the high flammability of ethanol. The resonator was linked to a function generator through a single connection. The lock-in amplifier, along with the other instruments, was interfaced with the control computer (Hewlett Packard 9000 Series 300) for data collection.



Figure 4: Acoustic Resonator Setup¹²

For Raman spectroscopy, the acoustic resonator was replaced with stainless steel cell equipped with a sapphire window, and a 785 nm laser source was mounted vertically onto the cell, as shown in Figure 5. The cell served as the Raman scattering collection point, and the laser was connected to a spectrometer and an excitation source. The spectrometer was interfaced with an external laptop computer for data acquisition and control. The distance of the laser excitation source was predetermined using Snell's Law.



Figure 5: Raman Spectrum Measurement Setup

The experimental schematic setup for both speed of sound measurement and Raman spectroscopy were based on the schematic from Trusler¹² as shown in Figure 6.

3.4 Experimental Procedure and Data Collection

The H₂ gas from a cylinder tank was directed through V1 and V2 to the reactor containing the activated catalyst at a steady increase in pressure until it reached approximately 8 MPa. During this flow, the reactor was immersed in a liquid nitrogen bath to shift the normal OP hydrogen composition away from its equilibrium value. Following catalyst exposure, V3 was shut, and the system pressure was increased up to 20 MPa using the manual hand pump. Downstream of V4, the H₂ flow was directed towards the acoustic resonator or the Raman measurement cell.

For both speed of sound and Raman spectrum measurements, the measurement data were collected using the corresponding software, a custom-written software and OceanView V2.0.20, respectively. The experiment conditions for each measurements are summarised as shown in Table 3 and Table 4.

Table 3: Summary of Experimental Conditions for Speed of Sound Measurements

Run	Pressure / MPa	Temperature / K
1	10	273.15
2 & 3	20	273.15

Table 4: Summary of Experimental Conditions for Raman Measurements

Run	Pressure / MPa	Integration Time / s
1	20	45
2	20	20

3.5 Data Analysis

3.5.1 REFPROP

To evaluate the accuracy and the reliability of the experiment results, REFPROP was used to provide first estimations of speed of sound data at different OP ratios and different pressures. The estimated results were then further analysed and used as a reliable reference point for experiment runs.⁷

3.5.2 Combined Conversion Rate Constant, k''

In general, OP conversion requires an inhomogeneous magnetic field which provides a magnetic field gradient. In the absence of an external catalyst, the conversion proceeds via a homogeneous autocatalytic mechanism.³

As para-H₂ is non-magnetic, the collisions between ortho-H₂ molecules and the local magnetic moments are considered to govern the homogeneous conversion

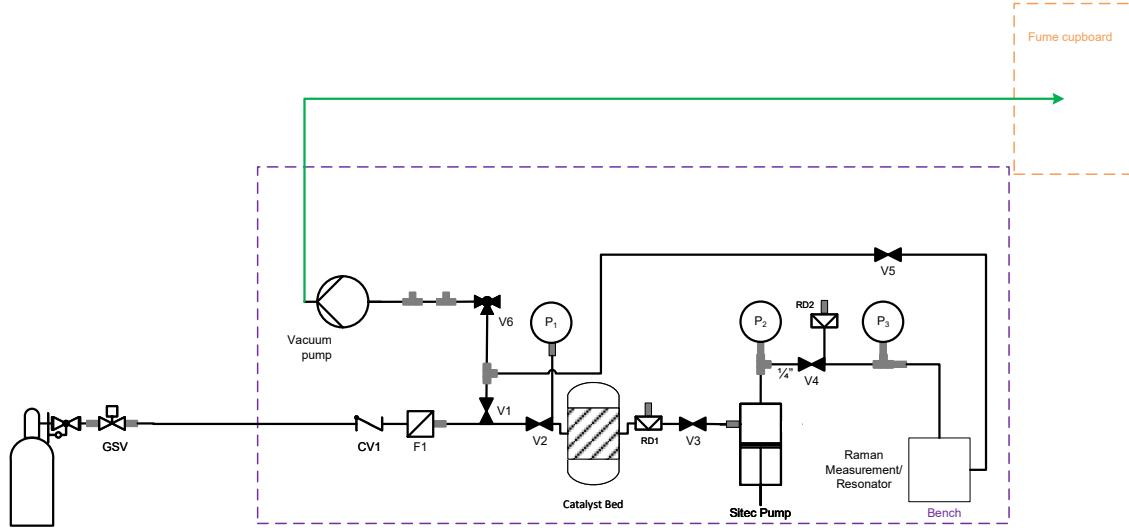


Figure 6: Process Flow Diagram of the Measurement System¹²

rate,³ which typically follow second-order kinetics as shown in Equation 10.¹³

x represents ortho- H_2 mole fraction, x_e represents the equilibrium ortho- H_2 mole fraction, k is the OP conversion rate constant, and k' is the PO conversion rate constant. k and k' can be correlated to a single combined conversion rate constant, k'' , as shown in Equation 11.

$$\frac{dx}{dt} = -kx^2 + k'x(1-x) \quad (10)$$

$$k = \frac{k'(1-x_e)}{x_e} = k''(1-x_e) \quad (11)$$

The solution of this second-order rate equation follows the general representation as shown in Equation 12. x_0 represents initial ortho- H_2 mole fraction and t represents time.

$$x(t) = \frac{x_0 x_e}{x_0 + (x_e - x_0) \exp(-k''t)} \quad (12)$$

As the speed of sound and ortho- H_2 mole fraction are linearly dependent, the same approach is used to estimate $c(t)$, speed of sound at any given time.

A scalar parameter, b , is added to account for the pressure drop that occurs in the system, and P is the real-time pressure measured in MPa. The term $P - P_0$, accounts for the pressure drop that occurred during the experiment runs, where P_0 is the designated experiment pressure in MPa, which was 10 MPa and 20 MPa, respectively.

$$c(t) = \frac{c_0 c_e}{c_0 + (c_e - c_0) \exp(-k''t)} + b(P - P_0) \quad (13)$$

Using Equation 12 and 13, the parameters are determined by minimising sum of squares of error (SUMSQ) obtained when fitting experimental data and the proposed rate equation for both analysis methods. As the experiment approaches equilibrium, the second-order rate equation will approach first-order behaviour.

3.5.3 Raman Spectrometer Data Analysis

3.5.3.1 Method of Determining OP Ratio

As mentioned in Section 2.5, the H_2 molecule is assumed to always be in ground vibrational state. This signifies that the population of ortho- H_2 and para- H_2 are only correlated to the rotational states, which follows a Boltzmann distribution.²

The Boltzmann distribution is calculated as a function of temperature, and this is used to derive Equation 14. $x_o(T)$ is the ortho-fraction of H_2 for a certain temperature T , L is the rotational quantum number, d_o and d_p are the degeneracy of ortho- H_2 and para- H_2 respectively, and F_L is the energy of the corresponding rotational state.²

$$x_o(T) = \frac{\sum_{L=1,3,5,\dots} d_o(L) e^{-F(L)/kT}}{\sum_{L=0,2,4,\dots} d_p(L) e^{-F(L)/kT} + \sum_{L=1,3,5,\dots} d_o(L) e^{-F(L)/kT}} \quad (14)$$

The contributions to the sum as shown in Equation 14 can be interpreted as the relative probability of an H_2 molecule, $J_L(T)$, to achieve a certain rotational quantum number L at temperature T , as shown in Equation 15.² The derived J values are dependent on temperature and are used for calibration and further determination of the actual OP ratio.

$$J_L(T) = d(L)e^{-F(L)/kT} \quad (15)$$

3.5.3.2 Calibration of Raman Analysis

The calibration approach applied here only refers to one reference measurement of H₂ with a known OP composition at a known Raman cell temperature. For this analysis, the calibration was conducted for normal H₂ at room temperature under a pressure of 20 MPa. This calibration method implements the J -level ratios determined as shown in Section 3.5.3.1.²

By correlating the J -level ratios, k_L , which are the individual calibration coefficients, can be determined for the respective H₂ isomer peak from single measurement. Based on the analysis conducted, only the first five peaks are easily distinguishable and were considered sufficient for OP ratio analysis. Hence, only L values ranging from 0 to 4 are investigated. Odd L levels are corresponding to ortho-H₂ population and even L levels are corresponding to para-H₂ population.

The set of simultaneous equations, as shown in Equation 16 to Equation 20 is solved to provide a unique set of k_L solution. A_L is the measured peak area of the individual isomer transition, which follows the relationship of $k_L \times A_L \propto J_L$.² A_L are calculated from the obtained spectrum by using the trapezoidal rule. f_{ortho} is the ortho-H₂ fraction calculated.

$$k_0 = \left(\frac{1 - f_{\text{ortho}}}{f_{\text{ortho}}} \right) \left(\frac{k_1 A_1 \times k_3 A_3}{1 + \frac{J_4}{J_0} + \frac{J_2}{J_0}} \right) \left(\frac{1}{A_0} \right) \quad (16)$$

$$k_1 = 1 \quad (17)$$

$$k_2 = \frac{J_2}{J_0} \frac{k_0 A_0}{A_2} \quad (18)$$

$$k_3 = \frac{J_3}{J_1} \frac{k_1 A_1}{A_3} \quad (19)$$

$$k_4 = \frac{J_4}{J_0} \frac{k_0 A_0}{A_4} \quad (20)$$

3.5.3.3 Actual OP Ratio Calculation

After calculating $k_L A_L$, the corrected areas were multiplied by J -correction as shown in Equation 21 and Equation 22.² These J -correction terms consider the contribution of each L level onto the final population contribution, which links the Boltzmann distribution to the actual area measured and used to calculate the actual OP ratio.

$$\frac{(k_1 A_1 + k_3 A_3) (J_1(T) + J_3(T))}{\sum_{i=1,3,5,\dots}^{\infty} J_i(T)} \quad (21)$$

$$\frac{(k_0 A_0 + k_2 A_2 + k_4 A_4) (J_0(T) + J_2(T) + J_4(T))}{\sum_{i=0,2,4,\dots}^{\infty} J_i(T)} \quad (22)$$

4 Results and Discussion

4.1 REFPROP Analysis

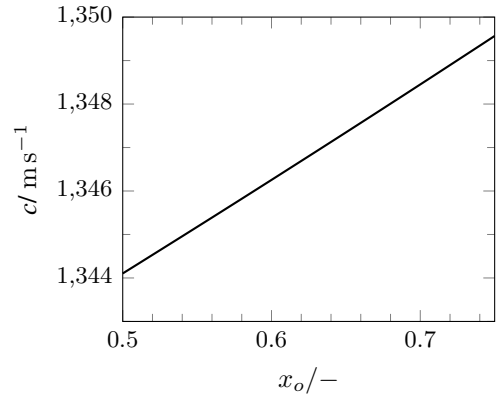


Figure 7: Speed of Sound against x_o at 10 MPa (REFPROP Analysis)

REFPROP⁷ is used to estimate the expected speed of sound against x_o at different pressures. Figure 7 portrays that the speed of sound at 10 MPa is linearly dependent on the x_o . Based on the analysis of REFPROP data, a linear dependence of speed of sound on x_o is observed at both 10 MPa and 20 MPa.

4.2 Speed of Sound Analysis

The results from the speed of sound revealed significant deviation when compared with the kinetics model of Milenko et al.¹³ The k'' values obtained by extrapolation beyond Milenko's original temperature limit of 120 K were found to be approximately twice the k'' values obtained for both different pressures.

While keeping parameters such as c_0 and c_e constant and excluding the pressure correction term b from the analysis, Milenko's model yielded higher speed of sound value at each iteration, indicating shorter time constant τ than initially expected as shown in Figure 8 and Figure 9.

The optimised parameters determined by using Equation 13 as discussed in Section 3.5.2 are shown in Table 5.

Table 5: Summary of Results Obtained from Speed of Sound Analysis

Run	1	2	3
c_0 / ms^{-1}	1344.25	1438.64	1438.11
c_e / ms^{-1}	1349.60	1441.96	1441.97
k'' / hr^{-1}	0.0070	0.0157	0.0182
Time constant, τ / day	5.95	2.65	2.29
$b / -$	9.226	9.253	8.956

Table 6: k'' Obtained From Extrapolation of Milenko et al.,

Run	1	2 & 3
k'' / hr^{-1}	0.0140	0.0263
Time constant, τ / day	2.98	1.58

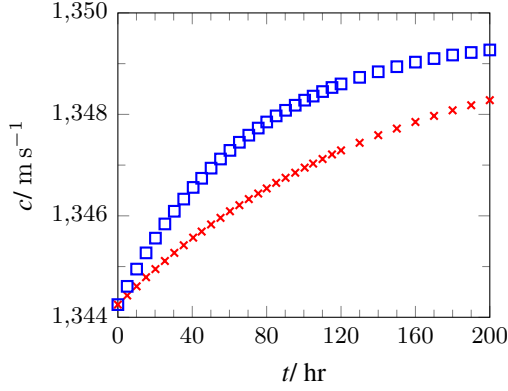


Figure 8: Comparison of OP Conversion Kinetic Model at 10 MPa

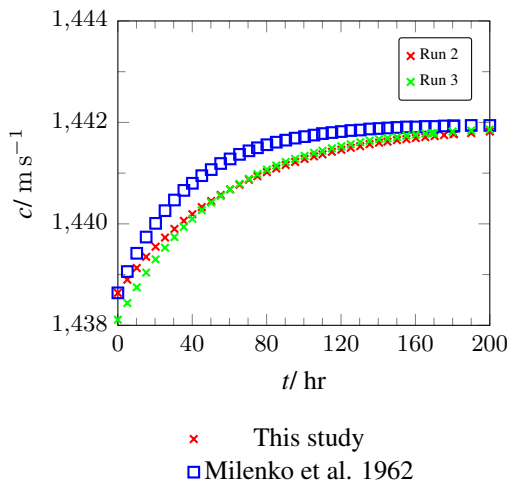


Figure 9: Comparison of OP Conversion Kinetic Model at 20 MPa

A substantial pressure drop was detected in the system, most likely caused by a leak due to misalignment of the PTFE ring at the resonator opening after prolonged use. Hence, measurement at 20 MPa was

repeated once (Run 3) to validate the result obtained from the initial run (Run 2). Although the change in the optimised parameters is insignificant, the k'' value in Run 3 was found to be roughly 15% higher than Run 2.

4.3 Raman Spectra

Figure 10 shows one of the Raman spectra obtained, and all 5 peaks are clearly distinguishable. The Raman spectrum shown in Figure 10 has had the background signals subtracted, which provides clearer peaks and a cleaner spectrum.

The background signals were presumably considered as a repeatable function of wavenumber arising from Raman scattering in an empty cell.¹² Hence, the background signal in this study was first acquired by obtaining the Raman spectrum of the system under vacuum. Raman shift shown in Figure 10 is defined as the frequency difference from the incident radiation, which is 785 nm in this study. The Raman shift is independent of incident radiation frequency, as it is a difference value.¹⁴

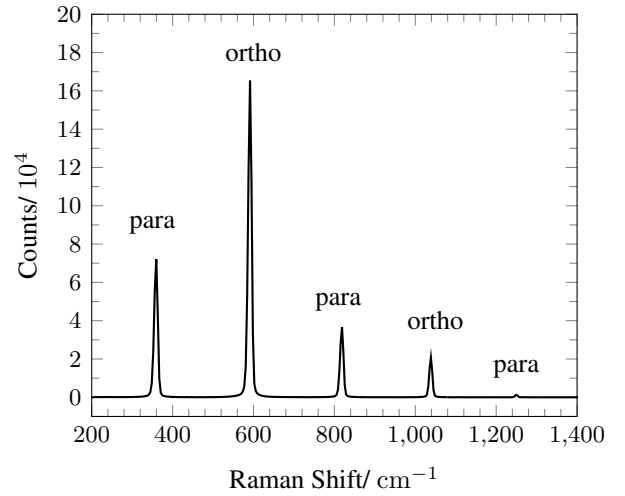


Figure 10: Raman Spectrum After Background Subtraction

Table 7 summarises the calibration coefficients determined using Equation 16 to Equation 20. Table 8 concluded the optimised parameters by considering Equation 12.

Table 7: Summary of Calibration Coefficients in Raman Spectrum 5 Peaks Analysis (After Background Subtraction)

k_0	k_1	k_2	k_3	k_4
0.619	1.000	1.124	1.221	1.314

Table 8: Summary of Results Obtained from Raman Spectrum Analysis

Run	1	2
x_0 / -	0.5187	0.5895
x_e / -	0.7480	0.7427
k'' / hr^{-1}	0.0275	0.0252
Time constant, τ / day	1.52	1.65

In the study, the accuracy of 4-peak and 5-peak analysis were considered alongside the significance of considering background noise elimination. To investigate the deviation between 4-peak and 5-peak analysis, the differences of the J -corrected orthofraction calculated between both analyses were considered. It was determined that the differences were at most at 0.02 %, which suggests an insignificant contribution by the fifth peak.

On the other hand, the significance of eliminating the background noise was also analysed with a similar approach. The differences were at most 0.08%, which also suggests that taking the background noise signal under vacuum is not necessary for determining the J -corrected orthofraction. However, a different combination of calibration coefficients needs to be determined for the analysis without considering vacuum elimination, as the calibration coefficients are a set of unique solution to a particular system.

In the study, the effects of J -correction terms as specified in Equation 21 and 22 onto the raw orthofraction calculated by considering only corrected areas were also considered. Based on the data obtained, the J -correction term has a significant correcting effect on the 4-peak analysis as compared to the 5-peak analysis.

A lower time constant is obtained from the Raman analysis, which does not align with the initial expectation, as it was conducted at higher temperature range. It was proposed that there was potential catalytic activity contributed by the stainless steel available in the Raman measurement setup.

4.4 Deviation From Proposed Milenko Model

The analysis of experimental data from speed of sound measurements and Raman spectroscopy has revealed deviations in the predictions of the Milenko model at 20 MPa when extrapolated above 120 K. As shown in Figure 11, the rate constant k'' values derived from the experiments differ from the model's extrapolated values. The speed of sound data shows a significant discrepancy, while data from Raman spectrum analysis is closely aligned with the model's predictions.

However, the validity of Milenko model extrapolated to ambient condition is yet to be confirmed. Hence, there is insufficient evidence that suggest Raman spectroscopy might capture molecular dynamics more accurately at elevated temperatures.

Based on the Milenko model, a peak was observed at approximately 170 K for the highest attainable k'' . It is proposed that this is due to the non-monotonic temperature dependence of the Milenko model. The k'' has been defined by a power function in the Milenko model, considering a single collision of two ortho- H_2 .¹³ There are also other competing effects affecting OP conversion suspected at high temperatures, such as 170 K and above. It is suggested that the rate of effective collision reduces at higher temperatures due to the shorter time of contact between ortho- H_2 molecules and the magnetic field, and hence contributes to lower k'' .

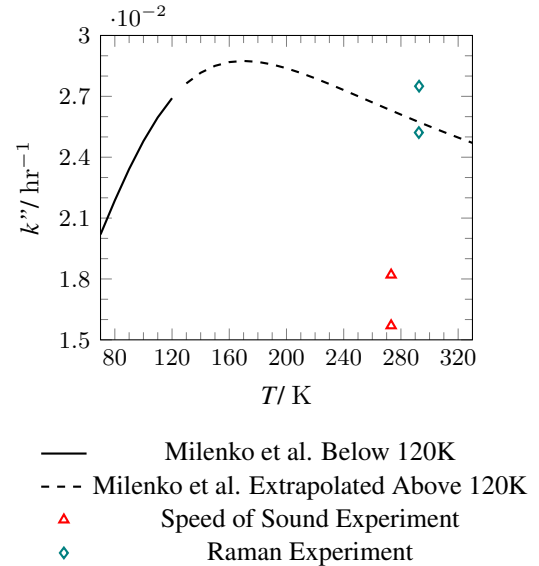


Figure 11: k'' Value against Temperature at 20 MPa

5 Conclusions

In this work, OP hydrogen conversion was characterised using speed of sound measurement and Raman spectroscopy. The stainless steel wall of the Raman cell may have potential catalytic activity for OP conversion, as opposed to the beryllium-aluminium alloy resonator wall. The surface area to volume ratio of the Raman cell is also higher than the resonator, which further enhances any surface-based catalytic activity. Thus, the significant difference in kinetic values obtained from these two methods might be attributed to the surface effects.

For future studies, one possible way to test the influ-

ence of surface catalysis would be to chemically passivate the stainless steel using an acid treatment and assess any changes in observed kinetics. Although the Raman-derived rate constants lie closer to the predictions of Milenko model at extrapolated temperatures, results from speed of sound measurement should not be disregarded since it is proven to be repeatable. Our present data do not allow a definitive validation of Milenko model above 120 K, and predictions at higher temperatures should be treated as a useful but unconfirmed reference.

6 Acknowledgement

We would like to thank Miriam Hauser for her constant support throughout this project.

References

- [1] Md Monjur Hossain Bhuiyan and Zahed Siddique. Hydrogen as an alternative fuel: A comprehensive review of challenges and opportunities in production, storage, and transportation. *International Journal of Hydrogen Energy*, 102:1026–1044, 2025.
- [2] Fabian Luther, Guinevere M. Sellner, Markus Richter, and Paul L. Stanwix. Inherent single-point calibration for quantitative raman analysis of ortho-para hydrogen composition. *International Journal of Hydrogen Energy*, 102:778–789, 2025.
- [3] Saif ZS. Al Ghafri, Stephanie Munro, Umberto Cardella, Thomas Funke, William Notardonato, J. P. Martin Trusler, Jacob Leachman, Roland Span, Shoji Kamiya, Garth Pearce, Adam Swanger, Elma Dorador Rodriguez, Paul Bajada, Fuyu Jiao, Kun Peng, Arman Siahvashi, Michael L. Johns, and Eric F. May. Hydrogen liquefaction: a review of the fundamental physics, engineering practice and future opportunities. *Energy Environ. Sci.*, 15:2690–2731, 2022.
- [4] Tan-Trieu-Giang Nguyen, Carsten Wedler, Sven Pohl, Dan Penn, Roland Span, J.P. Martin Trusler, and Monika Thol. Experimental speed-of-sound data and a fundamental equation of state for normal hydrogen optimized for flow measurements. *The Journal of Chemical Thermodynamics*, 198:107341, 2024.
- [5] Fuyu Jiao, Saif Al Ghafri, Keelan T. O'Neill, Paul S. Stanwix, Guinevere M. Sellner, Einar O. Fridjonsson, Neil Robinson, Eric F. May, and Michael L. Johns. Review of hydrogen ortho-para conversion: experimental data and reaction kinetics. *React. Chem. Eng.*, 9:2846–2862, 2024.
- [6] J.P.Martin Trusler. Final year research project 2025, 2025. PowerPoint presentation, Department of Chemical Engineering, Imperial College London.
- [7] Eric W. Lemmon, Ian H. Bell, Marcia L. Huber, and Mark O. McLinden. Reference fluid thermodynamic and transport properties (refprop) version 10.0. Software, NIST Standard Reference Database 23, 2018. National Institute of Standards and Technology (NIST), USA.
- [8] P. W. Atkins, Julio De Paula, and James Keeler. *Atkins' physical chemistry*. Oxford University Press, Oxford, twelfth edition / peter atkins, julio de paula, james keeler. edition, 2023 - 2023.
- [9] L. Ruffine and J.P.M Trusler. Sound-speed sensor for gas pipeline applications. *International Journal of Thermophysics*, 30:1106–1117, 2009.
- [10] Defne Demirdesen, Carsten Wedler, Robert E. Low, and J. P. Martin Trusler. Speed of sound in gaseous 1,1-difluoroethene (r1132a) at temperatures between 193 k and 383 k at pressures up to 3 mpa. *International Journal of Thermophysics*, 44(6):90, 2023.
- [11] Han Zhou, Zhengyu Li, Mengshi Li, Ming He, Qi Wu, and Linghui Gong. Study of activation methods for ortho-para hydrogen catalysts in a small isothermal converter based on gas chromatography at $\ln_2$ temperature. *International Journal of Hydrogen Energy*, 55:55–64, 2024.
- [12] J. P. Martin Trusler. Personal communication, 2025.
- [13] Yu. Ya. Milenko, R. M. Sibileva, and M. A. Strzhemechny. Natural ortho-para conversion rate in liquid and gaseous hydrogen. *Journal of Low Temperature Physics*, 107(1):77–92, 1997.
- [14] Q. Ye and P. Spencer. 9 - analyses of material-tissue interfaces by fourier transform infrared, raman spectroscopy, and chemometrics. In Paulette Spencer and Anil Misra, editors, *Material-Tissue Interfacial Phenomena*, pages 231–251. Woodhead Publishing, 2017.

Hybrid Normalised Advantage Function Reinforcement Learning for Energy-Efficient Switching Control in Temperature Swing Adsorption Systems

William Drury and Vilmos Turányi

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Temperature Swing Adsorption (TSA) processes are widely used in industry for gas drying and CO₂ capture but exhibit nonlinear switching dynamics between both adsorption and regeneration phases. The design of optimal control policies for these systems is challenging due to the characteristic coupled thermodynamic, kinetic and switching behaviour. Conventional reinforcement learning (RL) approaches often have difficulty handling hybrid discrete-continuous control structures efficiently in a single model, leading to excessively conservative switching regimes and inefficient energy usage. This work integrates hybrid normalised advantage value functions (HNAF) within an RL architecture to learn optimal switching and continuous control policies for TSA systems. The HNAF approach decomposes the hybrid action-value function analytically into a value term and a quadratic advantage term, enabling state mode selection and continuous control optimisation without numerical optimisation methods. A simplified TSA simulator incorporating a twin bed environment with heat and mass transfer dynamics was used to train the RL agent. The HNAF controller simultaneously learns the discrete mode (which bed is in adsorption or regeneration mode) and the continuous setpoints (mass flow rate and regeneration air temperature), forming an optimal policy for control. In comparison to a fixed-cycle baseline policy, the proposed HNAF-based hybrid RL policy reduced heating energy consumption by 29% and the number of bed switches by 40%, while adsorbing 32% less water. This resulted in a cumulative cost reduction of approximately 60%, and a normalised water-energy efficiency improvement of 59%, demonstrating that the HNAF framework is effective in optimising the trade-off between energy demand, switching frequency, and breakthrough penalties of TSA operation, as well as learning a robust policy for randomised initial states.

Introduction

TSA systems are vital for gas drying and CO₂ capture, yet prove challenging to control optimally, due to nonlinear, coupled heat-mass transfer dynamics and discrete switching between adsorption and regeneration. Conventional PID, model predictive control (MPC) and heuristic schedules struggle to optimally balance energy usage and cycle timing, while standard RL control methods can handle continuous or discrete actions but not both effectively. This is addressed by adopting a HNAF reinforcement learning formulation, decomposing the hybrid action-value into a mode dependent value and an analytic quadratic advantage term. This enables closed-form continuous actions and mode selection via advantage value comparison. Implemented in a simplified, lightweight twin-bed TSA simulation environment, the policy observes normalised state variables, and outputs both the operating mode and continuous variable setpoints (mass flow rate, regeneration air temperature) at each time step. This approach suggests a practical and generalisable framework for hybrid control in nonlinear switching chemical systems.

In this work, a dual-bed TSA operation is formulated as a hybrid Markov decision process (MDP) in which the RL agent must decide the optimal decision for both when to switch bed roles and how to adjust continuous setpoints. The HNAF architecture is adapted for this setting, providing analytic continuous minimisers for each feasible operating mode, which in turn avoids implicit numerical optimisation in the Bellman updates. A

simplified, but dynamically representative physical surrogate is developed for training, and is used for baseline comparison with the learned RL policy.

Background

This section summarises TSA operation and prior work on its control, before reviewing hybrid reinforcement learning methods for systems with mixed discrete-continuous actions. TSA systems are notoriously complex due to their nonlinear behaviour and coupled heat-mass transfer dynamics. Li et al. (2019) demonstrates the challenges of modelling and controlling such systems and emphasises that dedicated control strategies are required to handle these nonlinearities. Much of the existing TSA literature focuses on simulation and macro-optimal design rather than real-time control policies, for example in the process design and optimisation framework of S Krishnamurthy et al. (2024), which targets optimal equipment sizing and operating conditions rather than online decision-making. Where control is considered, existing approaches typically rely on single-loop controllers or relatively simple supervisory schemes. V. Skjervold et al. (2023), for instance, employs conventional feedback structures that struggle to capture the discrete-continuous dynamics associated with bed switching and regeneration. To date, these TSA control strategies do not treat the hybrid switching and continuous setpoint selection problem within a unified framework, and none have applied reinforcement learning to dual-bed TSA hybrid mode control.

Reinforcement learning is increasingly being utilised for process control due to its potential to adapt to changing operating conditions and to exploit complex nonlinear dynamics, as illustrated by recent studies such as N. Nievas et al. (2024) on RL-based optimisation. However, many RL formulations are designed for either purely discrete or purely continuous action spaces and handle hybrid control problems only with ad hoc constructions. M. Neunert et al. (2020) explicitly notes that “RL approaches have been optimised to work well with either discrete or continuous action spaces but can rarely handle both”, highlighting a key challenge for hybrid process control. Methods to address hybrid action spaces have therefore become an active topic for research. Fan et al. (2019) proposes a parameterised action space with separate actor networks for discrete and continuous decisions, which provides a flexible representation but can introduce interference between the two parts and does not yield an analytic optimum for the continuous parameters. Li et al. (2021) introduces an embedding-based method to represent hybrid actions in a lower-dimensional latent space, which improves scalability but reduces interpretability and weakens the direct physical meaning of control actions, both which are critical in safety and regulation-sensitive TSA applications. The HNAF framework proposed by Xiaofang Li (2023), which forms the basis of the present study, instead provides a unified analytic structure for hybrid actions by decomposing the hybrid action-value into a mode-dependent value function and a quadratic advantage term. This enables closed-form continuous action selection within each discrete mode and avoids the need the separate actor networks or opaque action embeddings, making it particularly well suited to the interpretable hybrid control required in TSA systems.

Methods

The objective is to operate a dual-bed TSA unit to minimise energy use and unnecessary switching while avoiding product breakthrough violations, subject to a minimum dwell-time between bed role changes.

Problem overview

This operation is modelled as a Markov decision process (MDP) with hybrid actions:

$$\mathcal{M} = (S, A, P, c, \gamma) \quad (1)$$

where S is the state space, A the hybrid space, P the transition kernel, c the cost and $\gamma \in (0,1)$ the discount factor.

The plant state $\mathbf{s}_t \in S$ at macro time step t collects the lumped inventories and temperature of the two beds (A and B) together with scheduling context:

$$\mathbf{s}_t = [x_{1,A,t}, T_{A,t}, x_{1,B,t}, T_{B,t}, ads_t, t_t^{dwell}, t] \quad (2)$$

where $x_{1,A,t}$ [kg] is the water inventory in each bed, $T_{A,t}$ [°C] is the bed temperature, $ads_t \in \{A, B\}$ indicates which bed is currently adsorbing, t_t^{dwell} is the time elapsed since the last role switch, and t is the current macro time.

The hybrid action, $\mathbf{a}_t \in A$ is decomposed as:

$$\mathbf{a}_t = (v_t, \mathbf{u}_t) \quad (3)$$

where $v_t \in \{0,1\}$ is a discrete decision and $\mathbf{u}_t$ is a continuous decision vector. The discrete variable encodes whether to keep the current bed roles ($v_t = 0$) or request a role switch ($v_t = 1$); a switch exchanges adsorption and regeneration roles between the two beds.



Figure 1. Illustration showing the decision variables for the TSA process in this study. The discrete binary decision v_t shown with either ‘stay’ or ‘switch’ for the bed role. The continuous decision setpoints $\mathbf{u}_t$ are shown with ‘Regeneration Temp.’ and ‘Mass Flow’ sliders.

A role switch is only feasible if the dwell-time constraint is satisfied:

$$t_t^{dwell} \geq t_{min}^{dwell} \quad (4)$$

where t_{min}^{dwell} is the minimum allowable dwell time between switches.

The continuous control $\mathbf{u}_t = [T_t^{reg}, F_t^{tot}]$ contains the regenerator inlet air temperature T_t^{reg} [°C] and the total inlet mass flowrate F_t^{tot} [kg/s] supplied to both beds. The adsorber inlet temperature is held constant for this study and is therefore not included in $\mathbf{u}_t$. Both T_t^{reg} and F_t^{tot} are subject to actuator bounds reflecting realistic operating constraints.

The next state is generated by a dual-bed TSA physics surrogate, written stochastically to allow for possible noise and disturbances:

$$\mathbf{s}_{t+1} = P(\cdot | \mathbf{s}_t, \mathbf{a}_t) \quad (5)$$

Generally, RL learning formulations tend to maximise a reward function, where good behaviour is assigned a conventionally positive contribution. In this study, an additive stage cost $c_t = c(\mathbf{s}_t, \mathbf{a}_t)$ is minimised for the purpose of implementing the HNAF structure.

$$c_t = \lambda_E \Delta E_t + \lambda_S 1\{\text{switch}_t\} + \lambda_Q \max(0, \dot{m}_t^{\text{out}} - \dot{m}_t^{\text{spec}}) - \lambda_\alpha \Delta m_t \quad (6)$$

$$\text{st. } \lambda_E, \lambda_S, \lambda_Q, \lambda_\alpha \geq 0$$

In this study, the stage cost c_t contains both positive contributions, which are penalties in this context, and negative contributions which are rewarded. The former comprises of the heater energy utilised (ΔE_t), switches present (switch_t), and the mass of water over the breakthrough limit ($\dot{m}_t^{\text{out}} - \dot{m}_t^{\text{spec}}$). The latter is the amount of water adsorbed in the stage (Δm_t).

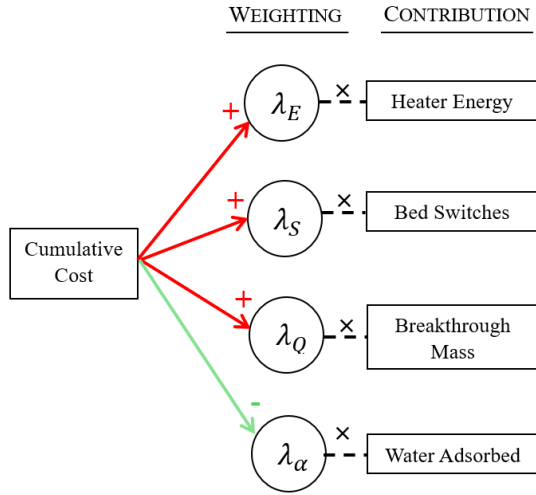


Figure 2. Diagram displaying the components of the cost, c , to be minimised. Positive contributions increasing cost, and therefore are penalised are shown in red, while the negative contribution, which is rewarded, is shown in green.

The optimal policy π^* minimises an expected cumulative cost:

$$\pi^* \in \operatorname{argmin} \left[\sum_{t=0}^{T-1} \gamma^t c(\mathbf{s}_t, \mathbf{a}_t) \right] \quad (7)$$

with a discount, $\gamma \in (0,1)$. Episodes run for a fixed horizon $T\Delta t$, with initial conditions $(x_{1,0}, T_0, \text{ads}_0)$ randomised within operationally sound ranges to promote policy robustness.

TSA Physical Environment

A dual-bed TSA physics surrogate is implemented and operated at discrete control intervals Δt . At any time t , one bed is in adsorption mode while the other is in regeneration mode. Each bed is represented by two lumped states (x_1^T) being the water inventory x_1 [kg] and the temperature of the bed T [°C]. The surrogate is simplified to be numerically stable when running on CPU only, while retaining the most influential mechanisms for control.

Saturated air is fed to the adsorber at temperature $T_{\text{ads},in}$ [°C] and a pressure P [bar]. The saturation pressure of water is computed via a

standard temperature dependant relation. The inlet humidity ratio w is then calculated based on saturation pressure and inlet pressure. The split of a commanded total mass flow $\dot{m}_{\text{tot}}$ into dry air $\dot{m}_{\text{air}}$ and water $\dot{m}_{w,in}$ follow:

$$\dot{m}_{\text{air}} = \frac{\dot{m}_{\text{tot}}}{1+w}, \quad \dot{m}_{w,in} = \frac{w \dot{m}_{\text{tot}}}{1+w} \quad (8)$$

Each bed also has a lumped heat capacity value:

$$mcp(x_1) = m_{\text{silica}} C_{p,bed} + x_1 C_{p,w} \quad (9)$$

Integration is carried out over micro-steps h with $\Delta t = Nh$ for numerical stability. Explicit Euler is used to update mass inventory for the TSA system while a semi-implicit linear update is utilised for temperature calculation.

For the bed in adsorption state, the water inventory increases via a smoothed capacity-limited rate, meaning that adsorption will slow down as the saturation of the bed increases.

$$x_1 = k' \dot{m}_{w,in} r_{\text{ads}}(x_1) \quad (10)$$

$$r_{\text{ads}}(x_1) = \frac{1}{2} \cdot \left[1 - \tanh \left(100 \left(\frac{x_1}{m_{\text{silica}}} - C_{\text{max}} \right) \right) \right] \quad (11)$$

The temperature update accounts for sensible gas heating or cooling, β_{sens} , and the isosteric heat term, β_{iso} :

$$T^+ = \frac{T + \frac{h}{mcp(x_1)} [\beta_{\text{sens}} T_{\text{ads},in} + \beta_{\text{iso}}]}{1 + \frac{h}{mcp(x_1)} \beta_{\text{sens}}} \quad (12)$$

$$\beta_{\text{sens}} = (\dot{m}_{\text{air}} C_{p,air} + \dot{m}_{w,in} C_{p,w}) \quad (13)$$

$$\beta_{\text{iso}} = \Delta H_{\text{ads}} \dot{x}_1 \quad (14)$$

Desorption in the TSA process is modelled by a linear driving force (LDF) type law with an Arrhenius temperature dependence and a mild inventory-shaping factor, $g(x_1)$, to slow the desorption when the water content of the bed is near empty:

$$\dot{x}_1 = -k_{\text{des}}(T) (x_1 - x_{\text{eq}}(T)) + g(x_1), \quad (15)$$

$$k_{\text{des}}(T) = k_0 \exp \left(-\frac{E_a}{RT_K} \right),$$

$$g(x_1) = \left(\frac{x_1}{m_{\text{silica}} C_{\text{max}}} \right)^{0.8} \quad (16)$$

The equilibrium inventory $x_{eq}(T)$ decays with temperature, and is calculated as a fraction of capacity:

$$x_{eq}(T) = (m_{silica} C_{max}) \cdot (x_{frac}^{20^\circ\text{C}} \exp[-k_T(T - 20)]) \quad (17)$$

The temperature update is calculated in the same way as it is for adsorption, with the dependence on $T_{ads,in}$ being replaced with $T_{reg,in}$, and:

$$\beta_{sens} = \dot{m}_{reg} C_{p,air} \quad (18)$$

To model breakthrough at the adsorber outlet a lightweight, calibrated surrogate, based on a moving mass-transfer zone (MTZ) with a logistic front, is used. This gives a realistic breakthrough curve without solving PDEs, for the purpose of calculation at RL time scales and makes the outlet response explicitly dependent on bed utilisation and gas flow. Let $\mathfrak{u}$ denote the current adsorber's fractional utilisation,

$$\mathfrak{u} \triangleq \frac{x_{1,ads}}{m_{silica} C_{max}} \in [0,1] \quad (19)$$

where $\mathfrak{u}$ is a measure of bed saturation (0 = fresh, 1 = at capacity). The outlet water mass flow follows:

$$F_{w,out}(\mathfrak{u}, F) = Y_{eq} + [\dot{m}_{w,in}(F) - Y_{eq}]Y(\mathfrak{u}; F) \quad (20)$$

where $Y_{eq} = \emptyset \dot{m}_{w,in}(F)$ encodes the target outlet dryness ($\emptyset = 0$ for ideally dry product). The dimensionless outlet ratio Y is taken by the MTZ surrogate as a logistic wave:

$$Y(\mathfrak{u}; F) = \frac{1}{1 + \exp\left(-k_{flow}(F) \frac{\mathfrak{u} - \mathfrak{u}_{50}}{\varepsilon}\right)} \quad (21)$$

$$\mathfrak{u}_{50} = 1 - \frac{w}{2} \quad (22)$$

The MTZ width fraction $w \in (0,1)$ can be adjusted, where a smaller value gives a narrower MTZ and a later breakthrough. The steepness scaling factor $\varepsilon = \mathcal{O}(0.1w)$ links the logistic slope to the MTZ width fraction. This MTZ surrogate helps RL by providing a smooth differentiable map, avoiding discontinuous that would potentially destabilise value learning. It is also fast to evaluate and control-relevant; by exposing the key levers of utilisation and flow, the agent can learn when to hold, pre-heat and switch.

Finally, the regeneration heat duty is approximated as sensible heating of the regenerator stream:

$$\dot{Q}_{heater} = \dot{m}_{reg} C_{p,air} (T_{reg,in} - T_{ads,in}) \quad (23)$$

HNAF Formulation

The HNAF approach proposed by Xiaofang Li, et al (2023) can be incorporated into the problem. In the following explanation, let $\mathbf{s}$ denote the normalised state. For the discrete branch in the hybrid action $\mathbf{a}$, two scalar value heads $\{V_0(s), V_1(s)\}$ correspond to $v = 0$ and $v = 1$. A binary feasibility mask $\mathcal{V}_{feas}(s) \subseteq \{0,1\}$ prevents any illegal switches in respect to the minimum dwell condition. For the continuous branch, for each discrete mode v , a mean value $\boldsymbol{\mu}_v(\mathbf{s}) \in \mathbb{R}^2$ is selected for $\mathbf{u} = [T_{reg}, F_{tot}]$. During the RL model training, a small Gaussian exploration is added, and at evaluation the model acts greedily with $\mathbf{u} = \boldsymbol{\mu}_v(\mathbf{s})$.

The HNAF method breaks down the RL ‘ Q function’ into a state dependent value term and a positive-definite quadratic advantage so that the continuous optimum is at the mean:

$$Q(\mathbf{s}, \mathbf{u}, v) = V_v(\mathbf{s}) + (\mathbf{u} - \boldsymbol{\mu}_v(\mathbf{s}))^T \mathbf{P}_v(\mathbf{s}) (\mathbf{u} - \boldsymbol{\mu}_v(\mathbf{s})) \quad (24)$$

where $\mathbf{P}_v(\mathbf{s}) \geq 0$ is a learned, state-dependent precision via a Cholesky factorisation to ensure positive semi-definiteness. Hence, the continuous minimiser is made explicit for fixed $(\mathbf{s}, v)$ with $\mathbf{u}^* = \boldsymbol{\mu}_v(\mathbf{s})$ and $Q(\mathbf{s}, \mathbf{u}^*, v) = V_v(\mathbf{s})$, eliminating a continuous minimisation in the Bellman target, and update with only the mode term V_v .

For action selection, first the dwell-time mask is applied to identify $\mathcal{V}_{feas}(\mathbf{s})$. Next the mode which minimises v^* is selected, where:

$$v^* \in \operatorname{argmin} V_v(\mathbf{s}) \quad (25)$$

The corresponding continuous set-points are sampled where $\mathbf{u} = \boldsymbol{\mu}_{v^*}(\mathbf{s}) + \epsilon$ with small Gaussian ϵ which is clipped and denormalised to the product of both T_{reg} and F_{tot} bounded actuation ranges. The action $\mathbf{a} = (v^*, \mathbf{u}^*)$ is applied to the TSA simulator, and the next state, stage cost c , step mask and logs are received.

Mini-batches sampled from a replay buffer $\mathcal{D}$ are used to train off-policy. Each sampled transition contains the normalised state $\mathbf{s}$, action $\mathbf{a}$, stage cost c , next state $\mathbf{s}'$, termination flag $d \in \{0,1\}$, and feasibility mask for the next step $mask' \in \{0,1\}^2$, which encodes the dwell-time constraint for $v' \in \{0,1\}$. The Bellman target is defined using only value heads at the next state, masking the infeasible modes:

$$y = c + \gamma(1 - d) \min_{v'} V_{tgt,v'}(\mathbf{s}') \quad (26)$$

The target requires no continuous optimisation meaning only a two-way minimisation across feasible v' is needed. In the rare case that both modes are infeasible, the minimum value of

$(V_{tgt,0}, V_{tgt,0})$ is chosen as a fallback to avoid NaN values, however this does not alter behaviour since action selection at runtime separately enforces mode feasibility. The mean squared temporal difference (TD) error over the batch, $\mathcal{D} \sim (\mathbf{s}, \mathbf{a}, c, \mathbf{s}', d, \text{mask}')$, is minimised:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathcal{D}}[(Q_{\theta}(\mathbf{s}, \mathbf{u}, v) - y)^2] \quad (27)$$

Adam with gradient-norm clipping is used for this, with a soft target update for stability. If the termination flag, $d = 1$ the bootstrap term is set to zero via $(1 - d)$.

Implementation

All simulations and RL experiments were implemented in Python. Numerical operations and array handling were performed using NumPy, while the HNAF agent was implemented in PyTorch. The dual-bed TSA environment was written with a lightweight Gym-style interface enabling convenient interaction with the RL agents.

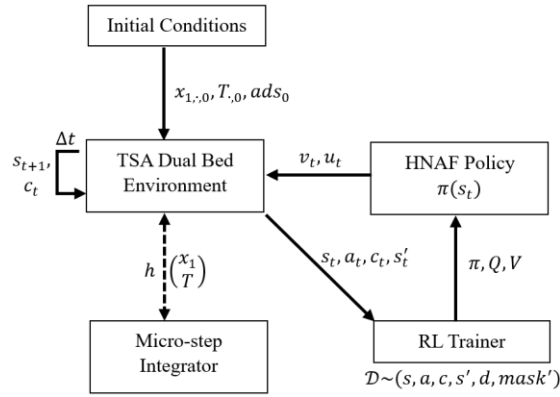


Figure 3. High-level block diagram of the proposed RL framework for dual-bed TSA operation. The HNAF agent provides hybrid actions $(v_t, \mathbf{u}_t)$ at each macro time step Δt . The TSA dual-bed environment performs micro-scale updates of the bed states over N steps of size h , with macro scale updates generating the next state $\mathbf{s}_{t+1}$ and stage cost c_t . Transitions are stored in a replay buffer and used by the RL trainer to update the policy and value functions.

The proposed RL framework in Figure 3. was deployed with a fixed-episode length of 32 hours, with macro time steps ΔT of 120 seconds and micro time step updates h of 30 seconds. This level of granularity enhances the accuracy and stability of the RL decision making process in periods around breakthrough or operational switches. It is also permissible due to the lightweight TSA surrogate environment utilised in training.

Table 1 RL training hyperparameters and objective weights used for experiments. Settings were chosen to ensure stable learning over long episode horizons and were held across fixed seeds unless stated otherwise. The cost weights represent the plant priorities of low thermal energy usage and switching, and breakthrough avoidance. Magnitudes are calibrated to realistic operating scales.

DEFAULT TRAINING VALUE		
Hyperparameter:		
Discount factor	γ_{tr}	0.9992
Learning rate	α_{tr}	3e-5
Hidden layers	$n_{h,tr}$	128
Batch size	$\mathcal{B}_{tr}$	128
Total steps	$n_{s,tr}$	2e5
Averaging rate	τ_{tr}	5e-3
Epsilon discrete	ε_{tr}	0.1
Noise std	σ_{tr}	0.15
Cost Weightings:		
Energy	λ_E	1e-4
Breakthrough	λ_Q	1e8
Switch	λ_s	200
Water adsorbed	λ_a	60

To establish a comparable baseline to assess the effectiveness of the proposed hybrid-RL solution, a conventional fixed-cycle approach is developed.

Table 2 Parameters for the baseline policy. The switch period was chosen to satisfy the breakthrough constraint under nominal feed flowrates and regeneration temperature. The continuous inlet setpoints match bounds available to the RL agent. The baseline considered is deterministic and non-adaptive.

BASELINE POLICY		
Parameter:		
Switch period (hr)	t_{switch}	6.0
Regeneration temperature (°C)	$T_{reg,in}$	210
Total inlet flowrate (kg/s)	F_{tot}	14.0

Results

This section presents the performance of the hybrid HNAF controller in terms of heater energy consumption, total water captured, number of bed switches and cumulative cost. Training stability and variability was first examined under randomised initial conditions in Figure 4., then the RL policy was compared to a fixed-cycle baseline using bed trajectory plots and aggregate metrics in Figure 5-6. Finally, sensitivity was analysed to training horizon and learning rate in Figure 7-8., and the trade-off between energy use and water recovery was summarised against trained policies in Figure 9.

During training, the HNAF agent interacted with the dual-bed TSA simulation environment using hyperparameters in Table 1. Initial bed inventories, temperatures and starting operational modes were randomised within operationally reasonable ranges to promote robustness to different starting conditions.

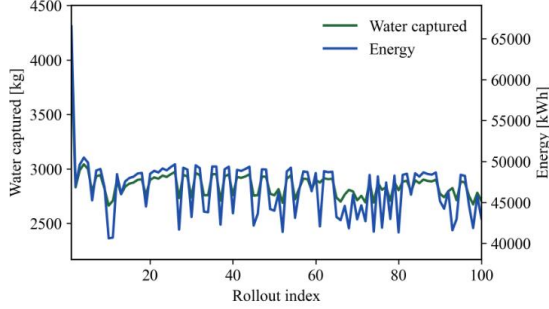


Figure 4. Per-episode key performance indicators for a single trained policy evaluated over multiple episodes with randomised initial states. Total water captured per episode (left axis, kg) and total heater energy usage (right axis, kWh) are plotted for each rollout in seed order.

Figure 4. presents the key performance indicators (total water captured and total energy consumption) for each episode based on the same learning strategy. The key performance indicators are high for the first rollout, which corresponds to the ideal initial conditions, after which both metrics cluster into a narrow band for the remaining initial conditions. This indicates that the policy converges to a consistent operating regime under varying initial conditions, and no numerically unstable or extremely high-cost trajectories were observed, meaning breakthrough events were extremely rare.

To assess the effectiveness of the proposed framework, the HNAF-based policy was compared with a conventional fixed-cycle baseline whose switch period, regeneration temperature and total inlet flowrate are given in Table 2. Figures 5-6. show the bed trajectories for the baseline and learned policies respectively, plotting bed water inventory (upper panels, kg) and bed temperature (lower panels, °C) for beds A and B, with vertical dashed lines indicating switches in bed roles. These time series illustrate how each policy schedules adsorption and regeneration over the 32-hour evaluation horizon.

The fixed-cycle baseline policy used 93650 kWh heater energy, executed 5 switches between the roles of the beds, and captured 5885 kg water during adsorption over the evaluation horizon. The cumulative cost associated with the process was 8.8e8. The HNAF-based RL-policy decreased energy consumption to 66498 kWh, while only executing 3 switches and capturing 4007 kg water concluding in a cumulative cost of 3.5e8.

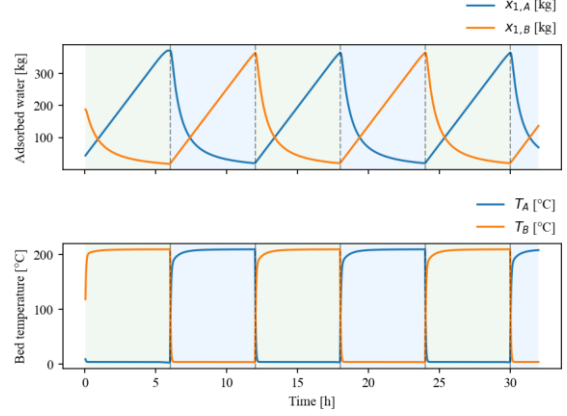


Figure 5. Bed trajectory plots for the baseline fixed-cycle policy using parameters from Table 2, showing adsorbed water (upper plot, kg) and bed temperature (lower plot, °C) for bed A and B respectively. Switching of bed roles can be observed by the vertical dashed lines.

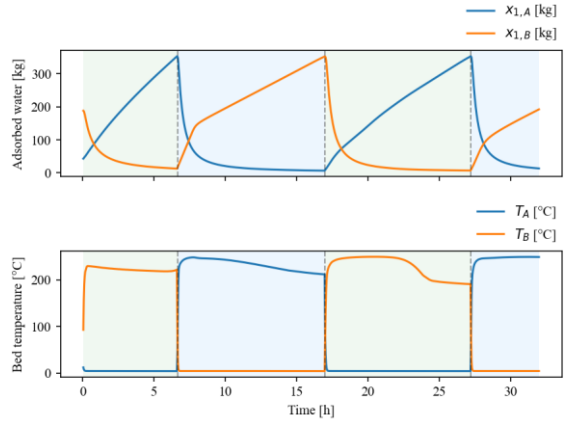


Figure 6. Bed trajectory plots for the leaned RL HNAF policy using hyperparameters in Table 1, showing adsorbed water (upper plot, kg) and bed temperature (lower plot, °C) for bed A and B respectively. Switching of bed roles can be observed by the vertical dashed lines.

Table 3 Key performance indicators for both the fixed-cycle baseline policy and the HNAF RL policy over a fixed 32-hour episode horizon. Constant seeded initial conditions were used for evaluation.

	BASELINE	HNAF
KPI:		
Water adsorbed (kg)	5885	4007
Heater energy (kWh)	93,650	66,498
No. of switches	5	3
Water-energy efficiency (kgH ₂ O kWh ⁻¹ switch ⁻¹)	12.6	20.1
Total cost	8.8e8	3.5e8

In relative terms, during the use of learned policy, energy consumption was decreased by 29%, switching events were reduced by 40% at the expense of decreasing captured water by 32%. With the specific weightings inside the cost function, this resulted in an approximately 60% of cost reduction. To compare how efficiently both the baseline and RL policy uses switching, a normalised water-energy efficiency (kgH₂O kWh⁻¹ switch⁻¹) can be

defined. The fixed-cycle baseline achieves $12.6 \text{ kgH}_2\text{O kWh}^{-1} \text{ switch}^{-1}$, while the learned HNAF RL policy reaches $20.1 \text{ kgH}_2\text{O kWh}^{-1} \text{ switch}^{-1}$.

Characteristic behaviour of HNAF-policy can be observed on the curves on Figure 5-6. Beds are kept in each mode for longer and only switched when the saturation is close to breakthrough zone. Temperature and flowrate in the regenerating bed are not kept high rather increased in critical stages. Adsorber outlet water content is allowed to approach closer to the specified limit with lower safety buffer than at baseline policy.

Sensitivity to standard RL hyperparameters is an important aspect of the practical validity of the HNAF framework. Training steps and learning rates were systematically varied for examining their effect on the learning.

During examination of training horizon, 10,000, 50,000, 100,000, and 500,000 environment steps were tested, and final evaluation costs were compared in Figure 7. Major improvement can be observed until reaching 100,000 steps. The policy quickly learns basic operating strategy (switching decision timing, regeneration efficiency levels) during this period. Negligible improvement was found during the further extension of training to higher number of steps.

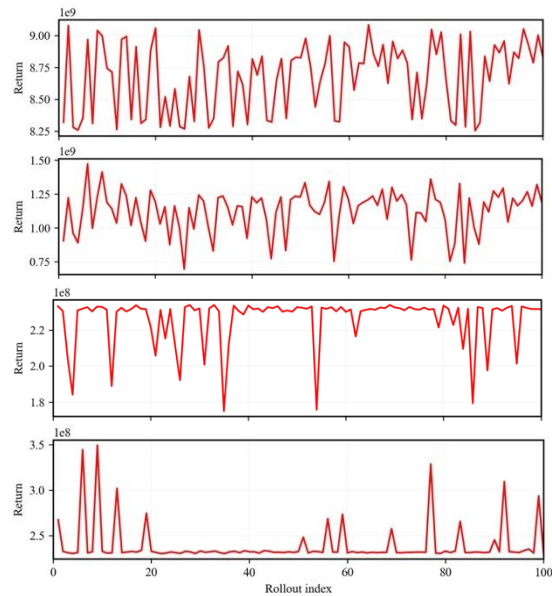


Figure 7. Effect of training horizon on cost return in a two-bed TSA system. Results of 100 rollouts with random initial states as a function of rollout-index are shown on each panel: policy trained for a) 10 000 steps, b) 50 000 steps, c) 100 000 steps, d) 500 000 steps.

Effect of learning rate was examined around the nominal value, $3e-5$ in Figure 8. Cost-return distributions are seen to be more concentrated, but at a higher plateau (worse cost) with low learning rates shown in Figure 8(a-b), and only traces of better rollouts can be found.

Curves converging to lower and more stable return levels are shown in Figure 8(c), from runs with intermediate learning rates. Moderate variance is observed between rollouts and most of the episodes staying at a narrow region. Efficient use of gradient information is indicated in this region while no outlier or divergent runs arise. Best evaluation costs were given by the nominal value seen in Table 1., which is in the intermediate region.

Episode returns are strongly scattered at high learning rates in Figure 8(d), with returns alternating between sharp peaks and deep outliers without conversion to a well-defined minimum, which points to the uncertainty and local instability of the learning rate.

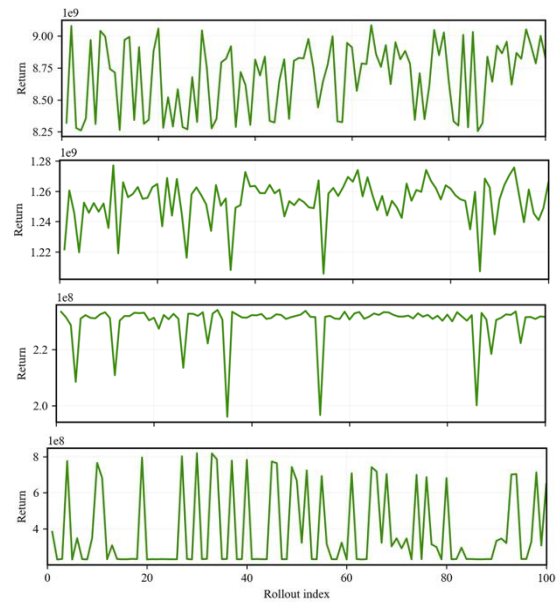


Figure 8. Effect of learning rate on cost return in a two-bed TSA system. Results of 100 rollouts with random initial states as a function of rollout-index are shown on each panel: policy trained with different learning rates a) $1e-6$, b) $5e-6$, c) $1e-5$, d) $1e-4$.

The results of the trained control policy in the energy–water plane is plotted in Figure 9. for a more detailed examination of the trade-off between process performance and energy use. Each point corresponds to the hybrid RL policy with different initialisations, the dominated solutions are represented by lighter markers, while the non-dominated, optimal solution set is represented by dark blue dots. The trade-off between energy consumed and water retained is clearly shown on the resulting front: rollouts with the lowest energy requirements generally remove less water, while achieving higher water yields requires significant additional energy. Diminishing returns are also indicated on the curve: beyond a certain water capacity, further improvement comes at increasing energy costs. In practical operation, an operating point could be chosen that provides noticeable energy savings compared to highly water-

maximising strategies, while only resulting in moderate water losses.

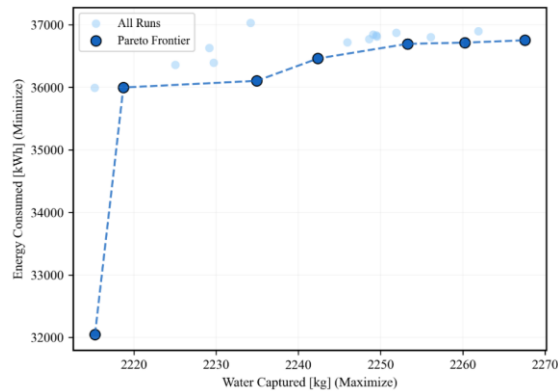


Figure 9. Optimal front for energy consumption and water retention for randomised initial conditions. All evaluated initial operating points are shown by the light blue dots, while the optimal solutions for which there is no alternative that would simultaneously reduce energy use and increase the amount of water captured are given by the dark blue markers, connected by the dashed line. The curve illustrates the compromises from which the operator can choose between more energy-efficient and higher water adsorption initial operating states are illustrated by the dashed line.

Discussion

The results of this study highlight multiple important insights regarding the behaviour and performance of using a HNAF RL controller for TSA systems. By comparing the learned policy to a fixed-cycle baseline and examining the policy evolution through different training parameters, it becomes clear that the RL agent exploited structural flexibilities of the dual-bed TSA environment in a way that legacy rule-based controllers can't. The following discussion evaluates these behaviours, linking them to the physical dynamics, RL decision mechanisms and the broader context of the challenges of hybrid control.

The predominant reason for improved performance is that the RL agent can dynamically adapt its decisions to the time dependent internal condition of each bed whereas the fixed-cycle policy relies on set switching intervals. This static behaviour inherently fails to account for variance in the thermal condition, regeneration efficiency, and the stochastic disturbances to the system. In contrast the RL agent monitors saturation levels, bed state temperatures, MTZ position and regeneration progress at each decision step, while forming a predicted advantage for switching. The agent repeatedly demonstrated that for an optimal approach, the beds can be kept in adsorption mode longer without causing breakthrough by delaying regeneration through continuous temperature actuation, until strictly necessary. This is shown in Figure 5-6. to reduce the total number of switches by 40%, showing that the baseline case over-performs regeneration relative to process needs, meaning that energy usage is not optimised.

An emergent behaviour is that the RL policy avoids maintaining high temperatures for the regeneration phase for the entire period. Instead, lower temperatures are utilised that applies heat energy strategically for periods when approaching critical desorption phases. The behaviour arises due to the HNAF formulation, where the RL agent is allowed to select continuous setpoints analytically, favouring values that minimise the advantage term while still acknowledging system constraints such as the breakthrough policy. The surrogate thermodynamic model of the TSA process means that heater has diminishing returns except near high saturation of the bed. The agent learned this through exploration and resultingly reduced total heat duty by 29% compared to the baseline case. Since the fixed-cycle policy continuously maintains a regeneration temperature setpoint of 210°C, a systematic energy waste is observed.

Besides temperature, the RL policy uses total inlet flowrate as a secondary optimisation decision variable. It can be observed in Figure 5-6. that flow is only increased in periods demanding rapid adsorption or assisted desorption. In most situations, lower flowrates are chosen to limit heating load and avoid driving the MTZ towards saturation levels. The fixed-cycle baseline has a constant inlet flowrate of 14 kg/s, ignoring the nonlinear dynamics of MTZ movement, and taking a more aggressive behaviour which accelerates breakthrough without improving regeneration efficiency to the same magnitude.

One of the most vital and important features of the learned RL behaviour is its willingness to operate close to the breakthrough limit. The MTZ surrogate model produces outlet wetness as a function of utilisation, allowing the trained controller to consider fine-grained differences between 85% and 95% utilisation. The RL agent effectively learns to 'squeeze' the system near breakthrough without premature switching to gain the maximum water adsorption in a safe way. This suggests that the policy can confidently interpret the MTZ dynamics and trusts that continuous controls can steer the bed away from unsafe behaviour near the breakthrough limit, improving overall efficiency with a small safety margin sacrifice. The overly conservative baseline switching, in contrast, unnecessarily sacrifices performance.

The RL policy is shown to capture ~32% less water than the baseline, however this is deliberate in the policy since water capture contributes far less to the cost function than the switching penalties and energy usage. With the training weights used in Table 1., the RL agent identifies that the limiting total adsorbed water reduces the regeneration frequency and heat input more than enough to compensate for the sacrificed adsorbed mass. This demonstrates a sophisticated

policy which emerges from the cost structure where the agent sacrifices some throughput to dramatically reduce the operational overhead, which ultimately led to a 60% lower cumulative cost than the baseline policy, and a 59% higher switch-normalised water-energy efficiency.

Typically, a challenge with RL in industrial systems is poor interpretability, especially with neural networks which must decide discrete and continuous action simultaneously. The HNAF architecture can mitigate this challenge through separate value heads for each discrete mode, and analytic minimisers for continuous actions. This makes the decisions of the agents understandable and transparent over black-box samples. This is reflected in the RL decision making behaviour, where mode switches occur when the mode dependent value indicates a higher long-term cost in staying. This interpretability also explains the stability observed in training and deployment of the model under different initial conditions, since mode-value comparisons are much easier to evaluate than learning a joint actor-critic mechanism on hybrid action space.

The influence of the total amount of training episodes shown in Figure 7. can be observed by the comparison of the cumulative cost achieved for greedy policy rollouts trained on varying total steps. The cumulative cost decreases by almost a magnitude for 10,000, 50,000, and 100,000 steps. This trend levels out for 500,000 demonstrating that a 200,000 step hyperparameter for the purpose of RL training is a suitable balance of computational speed and policy performance. The stability across the varying randomised initial conditions across the system for varying training step amounts suggests that the HNAF structure successfully prevented the common failures of RL in hybrid systems such as instability from joint discrete-continuous exploration, divergence due to non-smooth value surfaces, and forgetting due to the feasibility constraints applied to switching. Investigation into the learning-rate in Figure 8. confirms that intermediate learning rates produce both the best asymptotic cost and the most stable training dynamics. Low learning rates demonstrated policy instability under varying initial conditions and a high cumulative cost, while learning rates too high demonstrated high variance and oscillatory behaviour. This follows expectations from RL optimisation within hybrid systems. Since the Bellman target includes both masked discrete modes and precision matrices for the advantage term, carefully tuned gradient updates are still required to avoid drifting into unstable regions, despite the design reducing continuous optimisation difficulty.

Conclusions

The purpose of this study is to assess the feasibility of a HNAF RL control approach in the context of a TSA system. Several simplifications were contained in the problem formulation, which limits the direct transferability of the results to complex, industrial-scale TSA units. The function of TSA was defined as a hybrid Markov-decision process where the agent decides simultaneously on discrete mode (adsorption/regeneration switching) and continuous signals (regeneration temperature and total inlet flowrate). Heating energy and number of switches were significantly reduced at the cost of moderate reduction in adsorbed water in the simplified twin-bed model compared to the fixed-cycle baseline policy. Correspondent robustness was proven for initial states and hyperparameters. It is important to highlight that the computational requirements and number of samples used during learning exceed the tuning cost of a traditional controller, which may pose a significant obstacle in practical implementation.

The main contribution of this work is beyond numerical results. The HNAF-based RL frame handles the hybrid aspect of TSA control without the need for actor nets or difficult action embeddings. The HNAF architecture itself is not trivial, the per-mode covariance parameterization, the multi-node value function, and the handling of target networks entail significant model complexity, which can make maintenance and transparency difficult for industrial users. The quadratic advantage-structure is explicitly providing optimal interruption signals for each operating mode which simplifies the learning objective function and results in a more interpretable switching mechanism; the operating mode is decided by the value heads per mode while the setpoint values are directly provided by learning averages. This structure can be generalised and adapted to other cyclic or thermic joint processes where similar discrete-continuous connection is present. The method can only be used in a live production environment with additional layers of protection or in a supervised manner in its current form because it does not explicitly address regulatory constraints or hard security.

Moving to high-resolution process simulations and then to actual operational data from the current surrogate model is straightforward. This consists of coupling the HNAF-operator from the model side to detailed, non-equilibrium column-models (with axial dispersion, real-valve and actuator dynamics), and examining how stable the learned policy remains under measurement noise and model errors. One of the limitations of the present work is that the performance of the policy in the surrogate model was primarily assessed, so the risks arising from model-reality discrepancies have not been quantified yet. A next step is an 'advisory' implementation in a production environment, where

the RL agent recommends switching times and setting values that operators can accept or override, and runs in parallel with the existing DCS/MPC system. Gradual, secure validation of the regulator would be enabled and logged data for subsequent, off-policy fine-tuning would be provided. Such an advisory mode also assumes that operators trust the RL recommendations and understand their logic, which requires further explainability improvements and visualisation.

The integration of real-time process information and adaptive learning is an important further research direction. Adsorbent capacity, feed composition, equipment limitations, and contamination are constantly changing in practical TSA units. Therefore, the investigation of constrained RL formulations that explicitly handle safety limits (e.g., heating power limits) or employ safe online adaptation schemes, such as conservative gradient methods or policy updating based on periodically retrained surrogates could be included in future work. Currently, the discovery steps of online learning in the classical RL sense are difficult to reconcile with process safety expectations, therefore actual operational adaptation is expected to be permissible only in a highly limited, ‘safe’ action space. The controller, pre-trained on model ensembles to specific plants using a limited amount of plant data could be helped to transfer by domain randomisation and transfer learning.

The HNAF approach can be naturally extended to other swing or cyclic processes. These include fast cycle adsorption systems and pressure swing adsorption (PSA), distillation and membrane systems with periodic cleaning/regeneration steps, thermally integrated heat storage cycles, or certain batch and semi-batch reactors with discrete recipe steps. Similar discrete mode decisions and continuous operating targets are also featured in these applications, making the use of a unified, analytical RL controller an attractive option. In which cases HNAF-based RL offers a clear advantage over the MPC combination with traditional scheduling and can contribute to the development of generally reusable hybrid control software components and architectures in chemical industry practice can be clarified by system-level comparative studies on different processes.

Supplementary Information

Code available on request:

https://github.com/WMD159/tsa_rl.git

References

- Li, S.; Deng, S.; Zhao, L.; Xu, W.; Yuan, X.; Guo, Z.; and Du, Z. (2019) *Sci Rep* **9**, 16599.
- Krishnamurthy, S.; Cloete, S.; Pugno, V. (2024) *Ind. Eng. Chem. Res.* **64**, 12131 – 12144.
- Skjervold, V. T.; Nord, L. O. (2023) *Comput. Chem. Eng.* **178**, 108387.
- Nievas, N.; Pagès-Bernaus, A.; Bonada, F.; Echeverria, L.; and Domingo, X. (2024) *Appl. Artif. Intell.* **38**, 2383101.
- Neunert, M.; Abdolmaleki, A.; Wulfmeier, M.; Lampe, T.; Springenberg, J. T.; Hafner, R.; Romano, F.; Buchli, J.; Heess, N.; and Riedmiller, M. (2020) *arXiv preprint* arXiv:2001.00449.
- Fan, Z.; Su, R.; Zhang, W.; and Yu, Y. (2019) *arXiv preprint* arXiv:1903.01344.
- Li, B.; Tang, H.; Zheng, Y.; Hao, J.; Li, P.; Wang, Z.; Meng, Z.; and Wang, L. (2022) *arXiv preprint* arXiv:2109.05490.
- Li, X.; Dong, L.; Xue, L.; and Sun, C. (2023) *IEEE Trans. Neural Netw. Learn. Syst.* **34** (11), 9161–9170.

Inverse Engineering of Swift-Hohenberg Patterns via Physics-Informed Machine Learning

Zachariah Alan Mollatt, Maria Fernanda De Sousa Trad

Department of Chemical Engineering, Imperial College London

Abstract

Naturally occurring surface topographies are known to exhibit desirable functional properties, such as the hydrophobicity of a lotus leaf or the anisotropic wetting of rice leaves. Synthetically replicating these features by creating wrinkling instability is a promising avenue for creating functional microstructures. However, this remains challenging due to the extensive trial-and-error required to optimise multiple governing parameters that recreate these structures. This project poses an inverse engineering framework to accelerate the design-fabrication cycle by leveraging inverse Physics-Informed Machine Learning (PIML) to predict material and process parameters required to realise the desired functional patterns. We utilised the Swift-Hohenberg Equation (SHE) to model pattern evolution. We generated a 3D phase diagram that maps a governing bifurcation parameter ε to distinct morphological shapes including hexagons, stripes, and labyrinths. To solve the inverse problem, a Radial Basis Function Physics-Informed Neural Network (RBF-PINN) was developed to estimate ε from steady-state patterns. Model training was stabilised using three-phase curriculum learning, GradNorm adaptive weighting, and the decomposition of the fourth-order partial differential equation (PDE) into coupled second-order equations. The RBF-PINN results demonstrated high fidelity in parameter recovery, resulting in the computational reproduction of morphologies with low error. These digital patterns were experimentally validated by approximately replicating them onto polydimethylsiloxane (PDMS) substrates via soft lithography. Wettability tests confirmed that droplet footprint and surface coverage systematically vary with the underlying morphology, demonstrating that this integrated approach can effectively accelerate the engineering of functional surfaces.

1 Introduction

Patterned surface features, whether highly ordered or intrinsically disordered, are ubiquitous in nature, appearing across scales from microscopic ridges on butterfly wings to sand dunes and wrinkle patterns in soft materials [1]. These structures are not incidental; they reflect long evolutionary optimisation that equips natural surfaces with highly selective functional properties [1]. Leaf surfaces provide compelling examples: lotus leaves exhibit extreme hydrophobicity due to their pillar-like microstructures that give rise to the Cassie–Baxter state, as seen in Figure 1a-b, while the elongated ridges on rice leaves create anisotropic wetting (Figure 1c-d). Such observations highlight the potential of engineering synthetic surfaces with bio-inspired topographies to address industrial challenges such as antifouling, anti-icing, or self-cleaning surfaces.

In the literature, numerous strategies have been developed to generate surface patterns [1]. Among these strategies is the creation of functional surface microstructures by inducing controlled wrinkling instabilities. This provides a relatively simple and effective approach [2] which involves applying compressive strain to a bilayer elastomer film composed of a thin stiff (glassy) layer atop a compliant substrate, typically produced through UV oxidation or O_2 plasma treatment [2]. The resulting wrinkling instability arises from the competition between the film’s bending energy and the substrate’s elastic energy. However, engineering arbitrary surface morphologies through wrinkling remains challenging, as it requires optimisation over multiple parameters, including the viscoelastic properties of the

film, elastic modulus, viscosity, applied strain, film thickness, and the stress fields that collectively govern feature dimensions. This complexity often leads to extensive experimental trial-and-error. To overcome these limitations, we propose an inverse engineering framework based on PIML [3]. In this approach, a desired surface pattern is provided as input, and the PIML model predicts the material and process parameters required to realise it. By shifting the preliminary exploration from the laboratory to a digital environment, this strategy significantly reduces the engineering burden and accelerates the design–fabrication cycle. Our idea is to leverage this inverse mapping to rapidly identify the optimal conditions for producing targeted surface architectures, which can then be validated and fabricated experimentally.

In this work, we construct a synthetic digital topography by leveraging the physics of pattern formation [3], enabling us to generate tunable surface morphologies analogous to those found on natural leaves, such as pillars on a lotus leaf, and stripes on a rice leaf. These synthetic patterns serve as the target geometries for training a PINN, which learns the mapping between material parameters and the resulting surface features. The PINN predictions are first validated against the synthetically generated patterns and subsequently benchmarked using experimentally fabricated microstructures. Finally, we investigate droplet wettability on the patterned substrates and observe that both droplet footprint and surface coverage systematically vary with the underlying surface morphology. This integrated approach highlights how PIML can accelerate the design and engineering of functional textured surfaces.

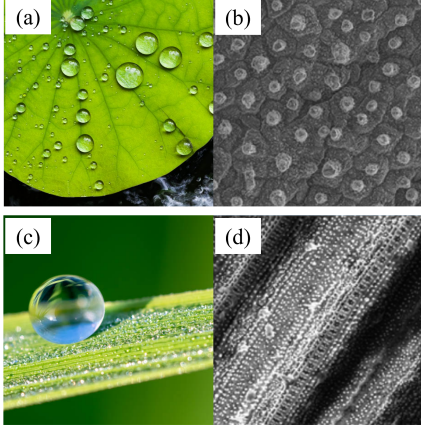


Figure 1: (a–b) Lotus leaf showing its hydrophobic behaviour, accompanied by an electron microscopy image of its surface topography. (c–d) Rice leaf demonstrating its wettability, along with the corresponding electron microscopy image illustrating its surface structure. [1, 4, 5, 6]

2 Theoretical Background

Surface wrinkles are a characteristic of mechanical instabilities that can develop over time in response to stresses and strains within a system. These morphologies gradually evolve toward a steady-state that minimises the overall free energy [7]. Mathematical models describing wrinkling instability capture the competition between bending and bending resistance using the Föppl-von Kármán equations, which are widely applied to model plate deformation [7]. Moreover, by incorporating a viscoelastic relaxation term into the Föppl-von Kármán framework, one obtains a PDE that provides a complete description of the system – capturing both the formation of wrinkle patterns and their evolution over time [8]. The PDE describing the evolution of the wrinkle pattern has a mathematical form similar to that of the Swift-Hohenberg Equation (SHE), which is widely used to model Turing patterns [8] as described below:

$$\frac{\partial u}{\partial t} = \varepsilon u - (\nabla^4 + k_c^2 \nabla^2 + k_c^4)u + N(u) \quad (1)$$

Here, $\partial u / \partial t$ describes the evolution of a field from a homogeneous unperturbed state to an approximated steady-state. In the case of wrinkling, $u(\mathbf{x}, t)$ is the out-of-plane displacement. The coefficient ε in the linear term represents the bifurcation parameter, commonly called the growth factor and k_c , the characteristic wavenumber, sets the intrinsic length scale of the resulting pattern, $\lambda = 2\pi/k_c$. The term $N(u)$ represents the nonlinearity arising from the stress-strain coupling, which governs pattern selection and saturation. In this paper, we choose a bistable nonlinearity with quadratic and cubic terms [9]:

$$N(u) = \delta u^2 - \gamma u^3. \quad (2)$$

Herein, the quadratic nonlinearity term, δu^2 , is responsible for breaking the reflection symmetry of the system,

$u \rightarrow -u$. This symmetry breaking promotes specific pattern morphologies. In contrast, the cubic nonlinearity term, γu^3 , ensures amplitude saturation and pattern stability. Consequently, the wrinkling pattern morphology, amplitude, and periodicity of the patterns are governed by three key parameters: ε , δ and γ . In the wrinkling scenario, ε is equivalent to the viscoelastic relaxation time [8].

3 Methodology

3.1 Numerical Solving Scheme

3.1.1 Generating Synthetic Wrinkled Patterns: Forward Problem

We numerically solve the SHE (shown in Eq.1) to generate the synthetic patterns resembling the wrinkle morphologies, yielding a time-dependent field, $u(\mathbf{x}, t)$. In principle, the SHE is stiff and non-linear. The stiffness of it arises from its fourth-order linear derivative, ∇^4 . In addition to the strong stability constraints imposed by PDEs, the calculation of fourth-order derivatives would be computationally expensive [10]. To overcome these challenges, a standard pseudo-spectral method with a semi-implicit time-stepping scheme was adopted [10, 11]. The pseudo-spectral method made use of the Fast Fourier Transform (FFT), which transforms the equation into Fourier space. The Fourier transform of the SHE is given by:

$$\frac{\partial \hat{u}(\mathbf{k}, t)}{\partial t} = \hat{L}(\mathbf{k})\hat{u}(\mathbf{k}, t) + \hat{N}(\mathbf{k}, t). \quad (3)$$

$$\hat{L}(\mathbf{k})\hat{u}(\mathbf{k}, t) := (\varepsilon - (k^2 - k_c^2)^2)\hat{u}(\mathbf{k}, t) \quad (4)$$

Where $\hat{u}$ is the two-dimensional Fourier transformation of the field u , and $\mathbf{k} = (k_x, k_y)$ is the wavevector, $\hat{N}(\hat{u})$ is the Fourier transformation of the nonlinear term, and $\hat{L}(\mathbf{k})$ is the transformation of the linear term. Whilst the Fourier transformation results in a simplification of the linear operator, applying this directly to the algebraic form of the non-linear term would result in convolution. Hence, a pseudo-spectral approach is used: the non-linear term is computed in real space, then transformed into Fourier space and combined with the Fourier-transformed equation. Note that while solving the complete SHE, the stiff linear part of the equation is treated implicitly. Non-linear terms are treated explicitly, or evaluated at the current time step, n . Thus, when the Fourier form of the SHE is discretised, it takes the following form:

$$\frac{\hat{u}^{n+1} - \hat{u}^n}{\Delta t} = \hat{L}(\hat{u}^{n+1}) + \hat{N}(\hat{u}^n). \quad (5)$$

This allows for the equation to be solved for the system's state at the next time step, $\hat{u}^{n+1}$, when using data calculated for the present time step, $\hat{u}^n$. For the purposes of this study, a characteristic wavenumber of $k_c = 1$ was selected. This parameter selection is common when modelling unbounded, infinite periodic domains [8]. The algorithm utilised for this is detailed below in Algorithm 1.

Algorithm 1: Solution to Swift–Hohenberg Equation

```

1: Input:  $\varepsilon, \delta, \gamma$ ; PDE coefficients,  $L_x, L_y$ ; grid lengths
    $N$ ; timestep  $\Delta t$ ; steady-state time  $t_{ss}$ ; tolerance.
2: Initialisation:
3:  $x \leftarrow \text{linspace}(0, L_x, N)$ ,  $y \leftarrow \text{linspace}(0, L_y, N)$ 
4:  $dx \leftarrow L_x/(N-1)$ ,  $dy \leftarrow L_y/(N-1)$ 
5:  $\text{Mesh} \leftarrow (dx, dy)$ 
6:  $k_x \leftarrow 2\pi \cdot \text{fft.freq}(N, dx)$ ,  $k_y \leftarrow 2\pi \cdot \text{fft.freq}(N, dy)$ 
7:  $k^2 \leftarrow k_x^2 + k_y^2$ ,  $\mathcal{L} \leftarrow \varepsilon - (1 - k^2)^2$ 
8:  $u \leftarrow \text{random uniform}(-\sqrt{\varepsilon}, \sqrt{\varepsilon}, (N, N))$ 
9:  $\hat{u} \leftarrow \mathcal{F}[u]$ 
10:  $u_{\text{prev}} \leftarrow u$ ,  $i \leftarrow 0$ ,  $\Delta_{\text{change}} \leftarrow \infty$ 
11: Time Evolution Loop:
12: repeat
13:    $i \leftarrow i + 1$ ,  $t \leftarrow i \cdot \Delta t$ 
14:    $u \leftarrow \mathcal{F}^{-1}[\hat{u}]$ 
15:    $\Delta_{\text{change}} \leftarrow \text{mean}(|u - u_{\text{prev}}|)/\Delta t$ 
16:    $u_{\text{prev}} \leftarrow u$ 
17:    $N(u) \leftarrow +\delta u^2 - \gamma u^3$ ,  $\hat{N} \leftarrow \mathcal{F}[N(u)]$ 
18:    $\hat{u} \leftarrow (\hat{u} + \Delta t \cdot \hat{N})/(1 - \Delta t \cdot \mathcal{L})$ 
19: until  $\Delta_{\text{change}} < t_{ss}$ 
20: Output:  $u_{ss} \leftarrow \mathcal{F}^{-1}[\hat{u}]$ 

```

3.1.2 Inverse Engineering: Identifying Parameters

The inverse problem for the SHE involves determining the unknown parameters:

$$p = \{\varepsilon, \delta, \gamma\} \quad (6)$$

from an observed steady-state pattern $u(x, y)$. Unlike the forward problem, where the solution field is calculated from known parameters, the inverse problem aims to recover the parameters through an inverse operator:

$$\mathcal{F}^{-1} : u_{\Omega} \mapsto p, \quad (7)$$

where Ω denotes the measurement domain. This approach follows the framework of Cho and Son [3], adapted for the SHE. Inverse problems for the SHE are challenging due to ill-posedness (multiple parameter sets may produce identical patterns), nonlinearity (creating a complex optimisation landscape), and fourth-order derivatives (amplifying sensitivity in parameter estimation).

PINNs are employed to estimate the parameters. PINNs integrate the governing PDEs directly into the network's loss function, allowing the network to approximate solutions that satisfy both observed data and the physical model. For the two-dimensional inverse problem, a PINN approximates the solution field $u(x, y)$ using a neural network $u_{\theta, p}(x, y)$, where θ are the network weights and p are trainable physical parameters. The loss function combines contributions from the data, the PDE, and, when necessary, boundary or initial conditions:

$$\min_{\theta, p} \mathcal{L}(\theta, p) = \mathcal{L}_{\text{Data}} + \mathcal{L}_{\text{PDE}} + \mathcal{L}_{\text{BC}} + \mathcal{L}_{\text{IC}}. \quad (8)$$

The data loss represents the difference between the predicted and measured data, the PDE loss enforces the SHE, the boundary condition loss enforces periodicity, and the total loss is minimised with respect to both θ and p :

$$\mathcal{L}_{\text{Data}}(\theta, p) = \frac{1}{N_d} \sum_{i=1}^{N_d} |u_{\theta}(x_i, y_i) - u^{\text{data}}(x_i, y_i)|^2, \quad (9)$$

$$\mathcal{L}_{\text{PDE}}(\theta, p) = \frac{1}{N_d} \sum_{i=1}^{N_d} |\varepsilon u - (1 + \nabla^2)^2 u + N(u)|^2, \quad (10)$$

$$\begin{aligned} \mathcal{L}_{\text{BC}}(\theta, p) = & \frac{1}{N_b} \sum_{i=1}^{N_b} |u_{\theta}(x_i, y_i) - u_{\theta}(x_i + L_x, y_i)|^2 \\ & + \frac{1}{N_b} \sum_{i=1}^{N_b} |u_{\theta}(x_i, y_i) - u_{\theta}(x_i, y_i + L_y)|^2. \end{aligned} \quad (11)$$

Numerical stability was improved by introducing an auxiliary variable to reduce the fourth-order PDE into two coupled second-order equations:

$$\mathcal{R}_1(u, p) = p - \nabla^2 u = 0, \quad (12)$$

$$\mathcal{R}_2(u, p) = \varepsilon u - (u + 2p + \nabla^2 p) + N(u). \quad (13)$$

Initial conditions are not required because the input data is assumed to be at steady state, and boundary conditions are imposed implicitly in the code. The adapted loss function sums the data loss and the two PDE losses:

$$\mathcal{L}(\theta, p) = \mathcal{L}_{\text{Data}} + \mathcal{L}_{\text{PDE1}} + \mathcal{L}_{\text{PDE2}}. \quad (14)$$

A three-phase curriculum learning strategy [12] is implemented to prevent convergence to local minima. The three phases gradually introduce loss terms and are defined respectively through the following equations:

$$\mathcal{L}_{\text{tot},1} = \mathcal{L}_{\text{Data}} \quad (15)$$

$$\mathcal{L}_{\text{tot},2} = w_1 \mathcal{L}_{\text{Data}} + w_2 \mathcal{L}_{\text{PDE1}} \quad (16)$$

$$\mathcal{L}_{\text{tot},3} = w_1 \mathcal{L}_{\text{Data}} + w_2 \mathcal{L}_{\text{PDE1}} + w_3 \mathcal{L}_{\text{PDE2}} \quad (17)$$

GradNorm adaptive weighting [13] is applied in phases 2 and 3 to control the relative magnitudes of the losses. For each loss, $\mathcal{L}_i$, GradNorm computes a desired gradient norm:

$$G_{\text{target},i} = \bar{G} \cdot \left(\frac{r_i}{\bar{r}}\right)^{\alpha}, \quad (18)$$

where $\bar{G}$ is the average gradient norm across all losses, α is a constant that controls the asymmetrical scaling of gradients, and r_i is the relative training rate of loss i :

$$r_i = \frac{\mathcal{L}_i(t)}{\text{moving average of recent } \mathcal{L}_i}. \quad (19)$$

If a loss decreases too rapidly, its corresponding weight, w_i , is reduced proportionally to maintain equilibrium. The moving-average approach stabilises training, particularly in

phase 3. Additionally, θ and $\mathbf{p}$ are optimised jointly with respect to the dynamically weighted loss function:

$$\mathcal{L}(\theta, p) = \sum_i w_i \mathcal{L}_i(\theta, p). \quad (20)$$

An RBF-PINN is used, adapted from Matas-Gil and Endres [14]. RBF networks consist of an input layer, a single hidden layer, and an output layer, and are effective for approximating smooth, continuous solutions to PDEs [15]. The general PINN framework can be seen in Figure 2 and the algorithm for parameter estimation can be seen in Algorithm 2.

Algorithm 2: Inverse RBF-PINN with GradNorm for Swift–Hohenberg Parameter Discovery

- 1: **Input:** $u_{\text{data}}, x, y, B, N_{\text{units}}, \sigma^2, \text{lr}_{\min}, \text{lr}_{\max}, \text{lr}_{\varepsilon,0}, \text{lr}_{\varepsilon,f}, \alpha, I_1, I_2, I_3$
- 2: Construct grid $(X, Y) = \text{meshgrid}(x, y)$; flatten U
- 3: Initialise trainable ε , fixed δ, γ
- 4: Initialise RBF-PINN $\mathcal{N}$ with N_{units} nodes; weights $w_d = w_1 = w_2 = 1$
- 5: Create Adam optimisers for $\mathcal{N}, \varepsilon$, and weights
- 6: Set iteration counter $i = 0$
- 7: **while** $i < I_3$ **do**
- 8: $i \leftarrow i + 1$; sample batch (x_b, y_b, u_b) of size B
- 9: Compute $(\hat{u}, \hat{p}) = \mathcal{N}(x_b, y_b)$ and Laplacians $\Delta \hat{u}, \Delta \hat{p}$
- 10: Residuals: $\mathcal{R}_1 = \hat{p} - \Delta \hat{u}, \mathcal{R}_2 = \varepsilon \hat{u} + \delta \hat{u}^2 - \gamma \hat{u}^3 - (\hat{u} + 2\hat{p} + \Delta \hat{p})$
- 11: Losses: $\mathcal{L}_d = \|\hat{u} - u_b\|^2, \mathcal{L}_1 = \|\mathcal{R}_1\|^2, \mathcal{L}_2 = \|\mathcal{R}_2\|^2$
- 12: **if** $i < I_1$ **then**
- 13: $\mathcal{L}_{\text{tot},1} = \mathcal{L}_d$; update $\mathcal{N}$
- 14: **else if** $i < I_2$ **then**
- 15: $\mathcal{L}_{\text{tot},2} = w_d \mathcal{L}_d + w_1 \mathcal{L}_1$
- 16: Compute G_d, G_1 ; periodically update w_d, w_1 via GradNorm; update $\mathcal{N}$
- 17: **else**
- 18: $\mathcal{L}_{\text{tot},3} = w_d \mathcal{L}_d + w_1 \mathcal{L}_1 + w_2 \mathcal{L}_2$
- 19: Compute G_d, G_1, G_2 ; update w_d, w_1, w_2 via GradNorm; update $\mathcal{N}$ and ε
- 20: **end if**
- 21: **end while**
- 22: **Output:** Estimated ε and trained RBF-PINN $\mathcal{N}$

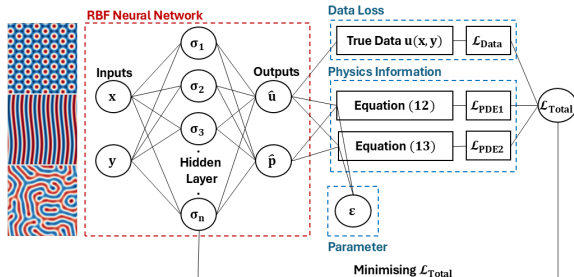


Figure 2: PINN architecture used for ε estimation.

3.2 Experiments

3.2.1 Scheme to Generate Wrinkled Patterns and Surface Replica

To investigate the wettability of SH-style patterns, the synthetically generated virtual patterns can be physically recreated onto PDMS surfaces. This process, commonly referred to as *soft lithography*, follows a three-step replication scheme based on standard literature guidance from Lama et al [2]. The steps are as follows:

1. **PDMS mould generation:** An inverse PDMS mould was prepared following standard soft lithography protocols, involving the mixing, degassing, and curing of Sylgard 184.
2. **Micro-wrinkle generation:** The PDMS mould was used to generate micro-wrinkles, which serve as the physical manifestation of the synthetic SH patterns.
3. **Surface replication:** These micro-wrinkles were replicated onto a fresh PDMS substrate, yielding the final patterned surface used for wettability measurements.

This scheme enables the study of wettability with varying surfactant surface tensions on the synthetically generated surface patterns.

3.2.2 Surface Wettability on Patterned Surfaces

A standard liquid was used for this investigation: water with Silwet L77 surfactant, adjusted to a surface tension of 30 mN/m. Liquid droplets of 2 mL were deposited onto each surface from directly above and allowed to spread. The spreading behaviour of the liquid across the patterned surfaces was observed and recorded to determine the influence of the underlying micro-patterned morphology on anisotropic wetting.

4 Results and Discussion

4.1 Pattern Generation and Phase Diagram

To construct a qualitative phase diagram, steady-state solutions to the SHE were generated and characterised across a three-dimensional parameter space defined by $(\varepsilon, \gamma, \delta) \in [0, 1]^3$. Each is classified based on its visual morphological pattern, yielding six distinct phase regions.

A large spatial resolution is used to ensure that the domain contains at least sixteen wavelengths. This is essential for fair classification; certain regimes may appear qualitatively different at smaller resolutions. For example, herringbone patterns may resemble simple stripes when observed on a smaller grid. Figure 3 depicts the three-dimensional phase diagram created. The phase diagram clarifies key pattern trends.

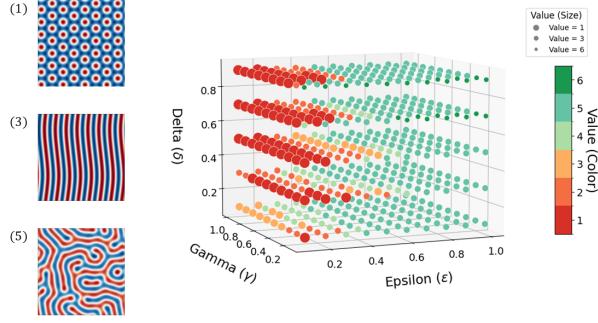


Figure 3: Phase diagram classification of SH patterns across $(\varepsilon, \gamma, \delta) \in [0, 1]^3$. Observed regimes: (1) hexagons, (2) mixed hexagons and herringbones/stripes, (3) herringbones/stripes, (4) mixed herringbones and labyrinths, (5) labyrinths, and (6) computational failure.

For weak initial instability (small ε) solutions are predominantly symmetric and exhibit highly organised patterns such as hexagons and stripes. As γ and δ increase, and the non-linear contributions grow relative to ε , hexagons begin to coexist with mixed morphologies. At this stage, patterns lose strict organisation but do not yet display full randomness, resulting in states where multiple structures coexist, such as hexagons with combinations of striped segments, or herringbone segments. When non-linear terms dominate, the solutions transition to highly disordered patterns. These include labyrinthine structures and irregular, randomised morphologies characteristic of strong non-linear interactions and anisotropic wrinkled patterns. This is due to the nature of the non-linear term, where δ is responsible for symmetry breaking, and γ is responsible for amplitude saturation.

4.2 Pattern classification

The classification of patterns was aided by a quantitative analysis of FFT graphs and Power Spectral Density (PSD) graphs, as seen in Figure 4.

FFT graphs reveal the underlying structure of the mor-

phology in the Fourier space. Hexagons exhibit six asymmetric peaks, whereas stripes are defined by two peaks. A labyrinth or a truly anisotropic pattern will display a characteristic isotropic central ring in the middle, representing the randomness of the structure.

PSD graphs provide a direct link to SH physics as the visually dominant peak will represent the characteristic wavenumber of the pattern, (k_c), and in turn the characteristic wavelength of the pattern (λ) [16]. In Figure 4, the wavenumber axis k is presented in grid units (pixel^{-1}); the dominant fundamental peak corresponds to the critical wavenumber $k_c = 1$ when scaled by the physical resolution of the simulation domain.

Additionally, the morphology of a patterned surface can be determined by specific features found in a PSD. Key identifiers include peak sharpness and distribution shape. Hexagons and stripes typically display a sharp, dominant peak in the 1D PSD plot, often followed by smaller harmonic peaks at higher wavenumbers. This signature reflects the strongly organised, long-range periodic nature of the patterns. Labyrinths display a broad base peak with no additional sharp harmonic peaks. The lack of peak sharpness reflects the disorder of the morphology, though a characteristic wavelength is still present. Hence, when converting the two-dimensional steady-state diagrams to PSD and FFT graphs, the visual classifications are consequently validated by the Fourier space.

4.3 Inverse engineering: Obtain parameters from the pattern

This section presents the inverse parameter estimation results using the RBF-PINN to recover the underlying parameter ε from observed steady-state patterns. The non-linear coefficients δ and γ were kept fixed. Three distinct patterns were analysed: hexagonal, stripe, and labyrinth. To ensure statistical robustness, each pattern was analysed across 10 independent simulations, utilising varying initial network seeds. Pattern accuracy was quantified using the Root Mean Squared Error (RMSE) and the Mean Squared Error (MSE).

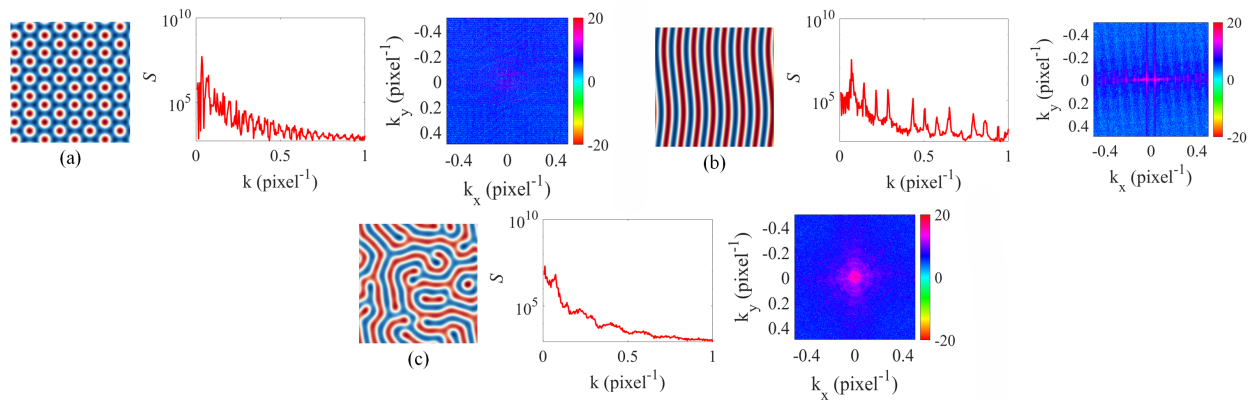


Figure 4: Spectral analysis of steady-state morphologies generated by the SHE for (a) Hexagons, (b) Stripes, and (c) Labyrinths. Left: Real-space patterns. Middle: 2D FFT (log scale). Right: Radially averaged 1D PSD (S).

4.3.1 FLOPS Time Complexity Analysis

The computational complexity of a single training iteration for the RBF-PINNs model can be estimated by a floating-point operations per second (FLOPS) calculation. In an RBF layer, one training iteration is defined by one forward pass, one backward pass, and a parameter update needed to calculate the second-order PDE residuals.

Hence, the computational cost per iteration is derived from the following components, where B represents the batch size and H denotes the number of RBF centers:

1. **RBF forward pass:** Calculating RBF activations by computing Euclidean distances from the B inputs to the H centers and then applying a kernel. Given an input dimension of D_{in} , this scales as $\mathcal{O}(B \cdot H \cdot D_{in})$.
2. **Dense layer forward pass:** Calculating the final output, with an output dimension of D_{out} , scales as $\mathcal{O}(B \cdot H \cdot D_{out})$ [17].
3. **Automatic Differentiation:** This step encapsulates the complexity of the backward pass for calculating the gradients required for parameter updates, as well as the nested differentiation needed for the calculation of second-order PDE residuals. Hence, it scales as $\mathcal{O}(B \cdot H \cdot D_{in})$ [18].

As the input and output dimensions ($D_{in} = 2, D_{out} = 2$) are small constant factors, they are disregarded in Big- $\mathcal{O}$ analysis. The dominant terms in complexity scaling are thus H and B . Hence, complexity scales linearly with batch size and number of nodes. The cost of a single fundamental step (one iteration) is dominated by the forward and backward passes over the batch:

$$\text{Time complexity per iteration} \approx \mathcal{O}(B \cdot H). \quad (21)$$

The total time complexity for the entire experiment must include the total number of iterations, I , because it represents the number of repetitions of the $\mathcal{O}(B \cdot H)$ operation. Unlike the spatial dimensions, I is a scaling parameter that directly affects total resource consumption in a non-constant way.

$$\text{Total Time complexity} \approx \mathcal{O}(I \cdot B \cdot H). \quad (22)$$

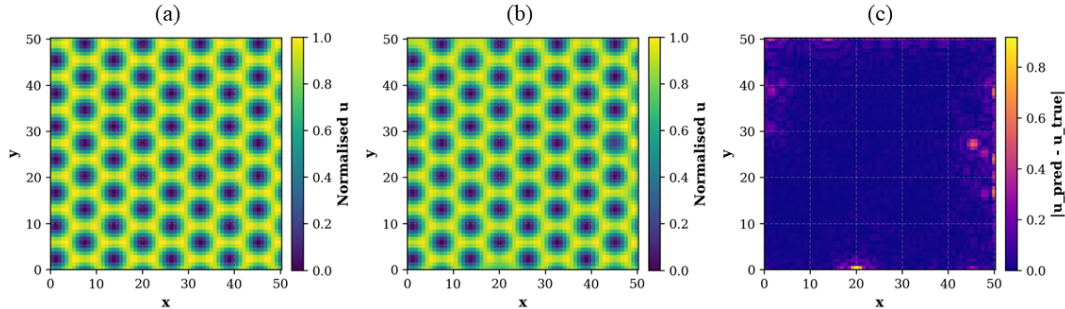


Figure 6: Comparison of the (a) true hexagonal pattern of a Swift-Hohenberg system and the (b) PINN prediction, with (c) absolute error showing the spatial difference between the two. The relative RMSE for the prediction is 5.64×10^{-2} .

4.3.2 PINN Hyperparameters and Sensitivity Analysis

A sensitivity analysis was conducted to find a hyperparameter configuration that balances computational cost with accuracy. Thirty grid searches were performed across several network sizes (200 to 1200 nodes) and batch sizes (16 to 256). This analysis focused on the system $\varepsilon_{true} = 0.6, \delta_{true} = 0.406$ and $\gamma_{true} = 0.196$, which defines a labyrinth morphology, representing the most complex pattern.

In Figure 5, the left panel demonstrates that larger networks and batch sizes lead to lower relative ε error, indicating improved parameter recovery when model capacity and gradient statistics are enhanced. The right panel (Giga-FLOPS) shows that the computation time increases with both the node count and the batch size. Based on this compromise, the final number of nodes was set at 300, and the final batch size was set to 256 samples.

Excessively large networks risk overfitting [19], due to the highly localised fitting of the discrete collocation points, breaking physical consistency. This results in $\mathcal{L}_{PDE}$ dominating $\mathcal{L}_{Data}$, leading to a non-physical ε . In contrast, a network that is too small cannot accurately represent the solution, and the reduced stability of the gradient prevents the reliable convergence of ε , producing an incorrect pattern.

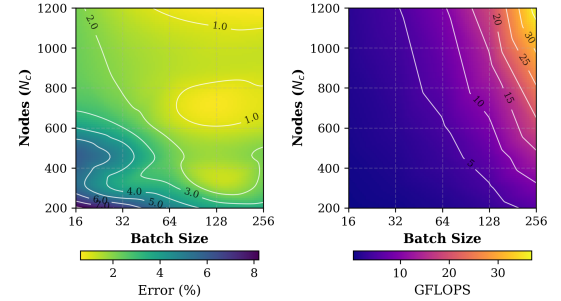


Figure 5: Performance-cost trade-off: left, interpolated relative ε error vs. network nodes and batch size; right, corresponding computational cost (Giga-FLOPS).

4.3.3 Hexagonal Pattern

Figure 6a shows the true hexagonal pattern generated from $(\varepsilon_{true}, \delta_{true}, \gamma_{true}) = (0.05, 0.406, 0.196)$. The PINN pre-

diction (Figure 6b) reproduces the structure of the pattern with high fidelity. The absolute error map (Figure 6c) exhibits uniformly low deviations across the domain.

The PINN reproduces the hexagonal morphology with low reconstruction error (RMSE $5.64 \times 10^{-2} \pm 1.41 \times 10^{-2}$, MSE $3.39 \times 10^{-3} \pm 1.54 \times 10^{-3}$). The small residuals indicate that the network captures both the geometric layout and the contrast of the hexagonal cells across the domain, rather than matching only local features.

The stability and physical consistency of the inverse process are confirmed in Figure 7, which shows a consistent decrease in all three loss components. The data loss was minimised the most, producing a pattern closely representative of the true solution. The final stabilised values were $\mathcal{L}_{\text{Data}} = 5 \times 10^{-3}$, $\mathcal{L}_{\text{PDE1}} = 8.1 \times 10^{-3}$, $\mathcal{L}_{\text{PDE2}} = 2.3 \times 10^{-2}$ demonstrating the efficacy of the neural network architecture.

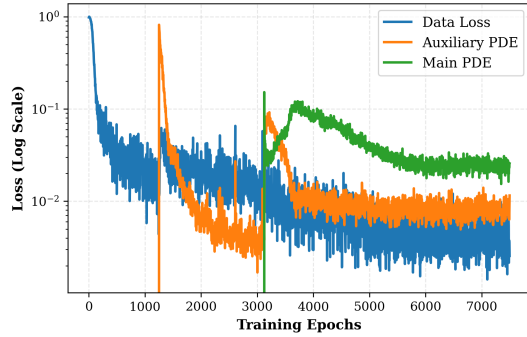


Figure 7: Log-scale loss convergence for hexagons.

The inferred value of ε was approximately 0.0812 ± 0.0213 , with a relative ε error of $62.4\% \pm 42.6\%$. This large difference arises because the true ε is near zero, where small variations in ε impact the pattern negligibly. Consequently, multiple ε values can produce similar patterns, allowing the PINN to converge to slightly different values while still reproducing the observed structure.

Figure 8 shows the convergence of the PINN prediction during training. The value remains near $\varepsilon \approx 0.5$ during phases 1 and 2, then rapidly decreases in phase 3, stabilising around 0.0812 after approximately 7000 epochs. The relatively large standard deviation reflects this non-uniqueness near the linear stability boundary.

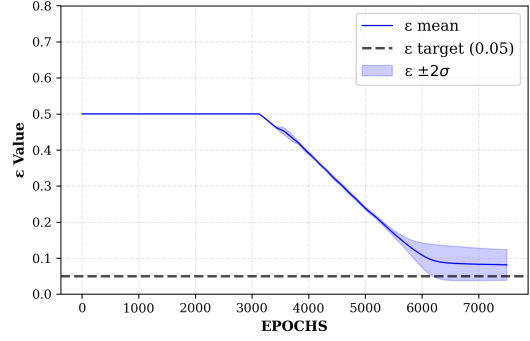


Figure 8: Epsilon convergence for hexagonal pattern.

4.3.4 Stripe Pattern

The stripe pattern generated with true parameters $(\varepsilon_{\text{true}}, \delta_{\text{true}}, \gamma_{\text{true}}) = (0.1, 0, 0.8)$ (Figure 9a) is accurately reconstructed by the PINN (Figure 9b), with minimal spatial deviation in the error map (Figure 9c). Low reconstruction errors (RMSE $= 8.14 \times 10^{-3} \pm 2.1 \times 10^{-3}$, MSE $= 7.08 \times 10^{-5} \pm 3.5 \times 10^{-5}$) demonstrate strong pattern similarity across the domain. The PDE simplifies with $\delta = 0$, removing the asymmetrical term and facilitating accurate PINN prediction of this linear, periodic pattern.

Among the three patterns, the stripe pattern achieved the lowest final losses (Figure 10), stabilising at $\mathcal{L}_{\text{Data}} = 8.2 \times 10^{-4}$, $\mathcal{L}_{\text{PDE1}} = 8.1 \times 10^{-4}$, and $\mathcal{L}_{\text{PDE2}} = 7.8 \times 10^{-4}$. The low residuals indicate the network accurately fits the data and satisfies the governing equations, reflecting the simplicity of this linear, periodic pattern.

The predicted ε has a mean of 0.1147 ± 0.007 , with a relative difference of $14.7\% \pm 7\%$. Although this difference may seem large numerically, small variations in ε near zero produce structurally similar patterns. Figure 3, verifies that similar patterns occupy discrete regions, explaining why the inverse recovery shows some variation despite accurate structural matching.

Figure 11 shows the convergence of ε . From phase 3, starting at 3100 epochs, ε steadily approaches the true value and stabilises around 0.1147. The narrow two-standard-deviation band indicates high reproducibility and precision confirming stable parameter recovery for the stripe regime.

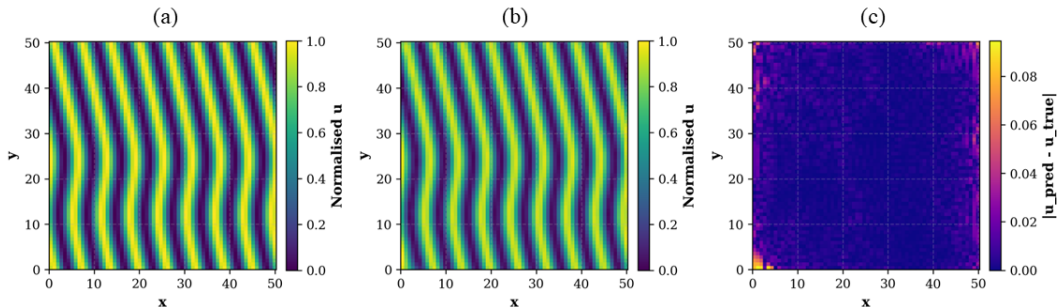


Figure 9: Comparison of the (a) true Stripe pattern of a Swift-Hohenberg system and the (b) PINN prediction, with (c) absolute error showing the spatial difference between the two. The relative RMSE for the prediction is 8.14×10^{-3} .

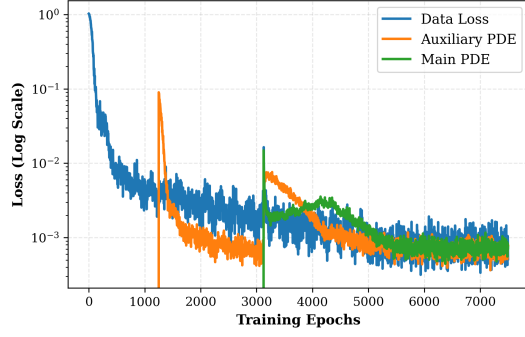


Figure 10: Log-scale loss convergence for stripes.

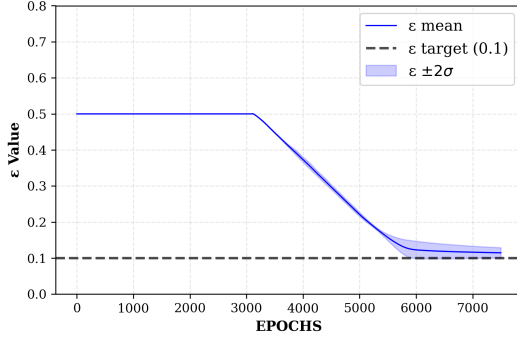


Figure 11: Epsilon convergence for striped pattern.

4.3.5 Labyrinth Pattern

The pattern produced using $(\varepsilon_{\text{true}}, \delta_{\text{true}}, \gamma_{\text{true}}) = (0.6, 0.406, 0.196)$ exhibits the characteristic labyrinth morphology (Figure 12a). The corresponding PINN reconstruction (Figure 12b) closely matches the spatial structure, with only small deviations visible in the error map (Figure 12c).

Reconstruction losses are higher than those for the hexagonal and stripe cases ($\text{RMSE} = 1.41 \times 10^{-1} \pm 3.3 \times 10^{-2}$, $\text{MSE} = 2.12 \times 10^{-2} \pm 9.44 \times 10^{-3}$) as the inherent randomness of labyrinth patterns makes pixel-level data fitting less precise.

The training losses (Figure 13) stabilise at $\mathcal{L}_{\text{Data}} \approx 2.2 \times 10^{-3}$, $\mathcal{L}_{\text{PDE1}} \approx 3.1 \times 10^{-2}$, and $\mathcal{L}_{\text{PDE2}} \approx 2.3 \times 10^{-1}$. The relatively higher $\mathcal{L}_{\text{PDE2}}$ is a consequence of the Grad-Norm adaptive weighting scheme applied in phase 3 (Section 3.1.2). This larger residual is expected for labyrinth

patterns due to their disordered structure and does not indicate network instability.

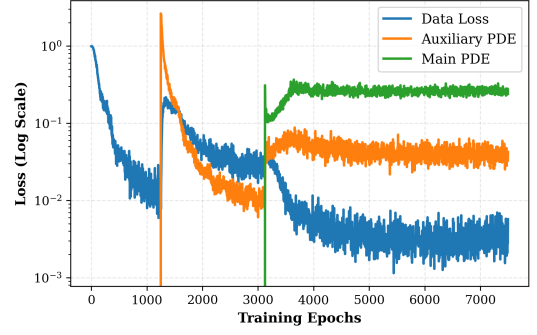


Figure 13: Log-scale loss convergence for labyrinth.

The recovered values of ε for the labyrinth pattern are both accurate and consistent averaging at 0.5849 ± 0.003 , with a relative error of $2.51\% \pm 0.5\%$, deviating only slightly from the true value. Figure 14 shows the convergence during training: in phase 3 (from 3100 epochs), ε initially decreases, then rises toward the true value eventually stabilising around 0.5849. The narrow two-standard-deviation band once again demonstrates the reproducibility and precision of the model, confirming that the PINN reliably converges to physically meaningful parameters despite the disordered nature of the labyrinth pattern.

Across all pattern types, the RBF-PINN reliably predicted ε values that accurately constructed the morphologies, but performance depended on how sensitive each regime is to parameter changes. Hexagonal and striped patterns, which form near the onset of instability, showed higher errors because small variations in ε produce patterns that look almost identical. In contrast, labyrinth patterns occur further from this boundary, so changes in ε lead to noticeably different morphologies, resulting in more accurate ε recovery. Overall, the method remains robust across regimes, with improved identifiability in the more non-linear labyrinth region. Slightly higher errors were observed near the domain boundaries for all patterns, suggesting that the implicitly applied boundary conditions may not fully capture edge effects. Addressing these boundary effects could further improve the accuracy of the PINN reconstructions.

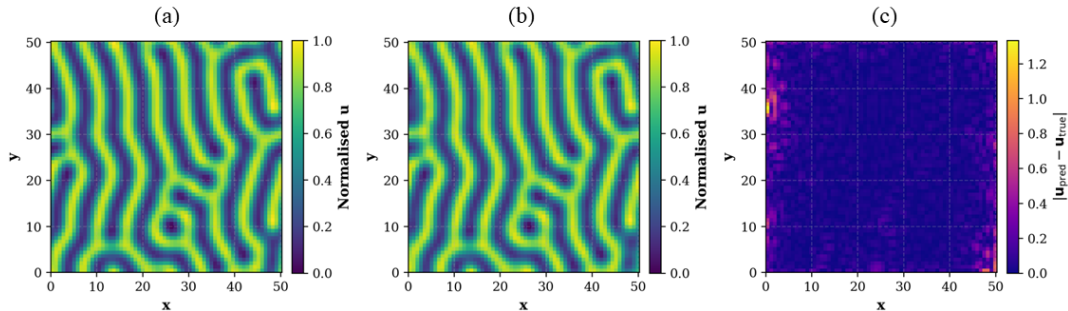


Figure 12: Comparison of the (a) true Labyrinth pattern of a Swift-Hohenberg system and the (b) PINN prediction, with (c) absolute error showing the spatial difference between the two. The relative RMSE for the prediction is 1.41×10^{-1} .

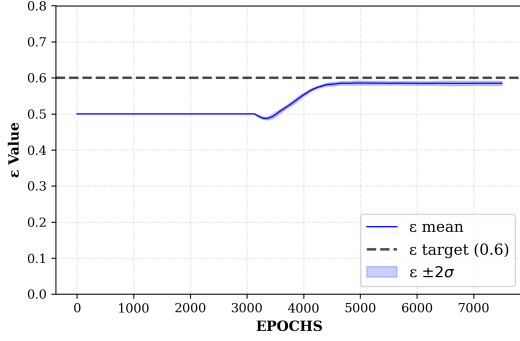


Figure 14: Epsilon convergence for labyrinth pattern.

4.3.6 Parameter Recovery from Experimental Wrinkled Patterns

The inverse-recovery framework was tested on an experimental wrinkled pattern formed on a PDMS substrate, as seen in Figure 15a. The image was cropped and down-sampled to 64×64 to suppress camera-induced edge shadows while retaining the core spatial structure.

The PINN estimated the governing parameters as: $\varepsilon = 0.2897$, $\delta = 0$, $\gamma = 0.0164$. These values were then passed through the forward problem architecture to generate the corresponding steady-state pattern shown in Figure 15b.

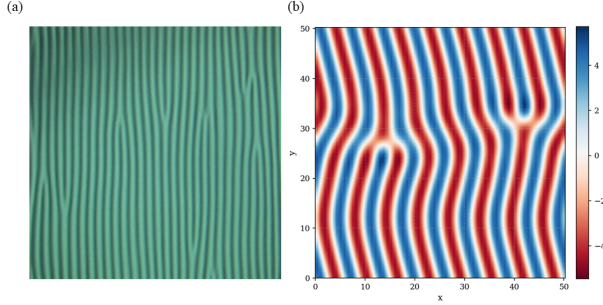


Figure 15: (a) Experimental wrinkle pattern on PDMS. (b) Pattern reconstructed from the recovered parameters.

The reconstruction reproduces the correct stripe morphology and captures the main defect structures, such as the cross-junctions along the lines. The main problem lies in the wavelength: the generated pattern exhibits a noticeably different peak spacing from the experimental image. This discrepancy highlights the limitations of the current model and indicates where refinement is needed in the inverse-recovery architecture.

This may be due to the simplification in the model's physics, where $k_0 = 1$ was assumed. As the observed pattern has a wavelength different from that of the generated patterns, the corresponding characteristic wavenumber also changes. In practice, the input image either needs to be preprocessed to match a target wavelength, or the PINN's physics must be adjusted to reflect the wavelength present in the data. Contrasting to the original generated patterns, the model actually predicted all 3 parameters, indicating that the current architecture is actually sufficient for the param-

eter recover for all three parameters. However δ and γ require further investigation and refinement for better results.

4.4 Wettability Test with the Desired Pattern

Figure 16 shows the wetting behaviour of the test liquid on four different surface patterns created on PDMS following the methodology in Section 3.2.1. Figure 16a is the control case on a smooth, unwrinkled substrate: the droplet forms an almost perfect circular shape, confirming isotropic wetting.

Figure 16b shows the droplet on a randomly wrinkled pattern. The droplet shape deviates from circular due to the surface undulations. Figure 16c corresponds to a herringbone pattern. Here, the contact line deformation is stronger and clearly anisotropic, demonstrating that the droplet spreads more easily along the dominant wrinkle direction than across it.

Figure 16d shows the droplet on a unidirectional stripe pattern. The droplet elongates noticeably along the stripe axis, while spreading minimally perpendicular to the stripes, confirming the pronounced anisotropic wetting driven by directional surface energy variations reminiscent to the functional surface properties observed on a rice leaf.

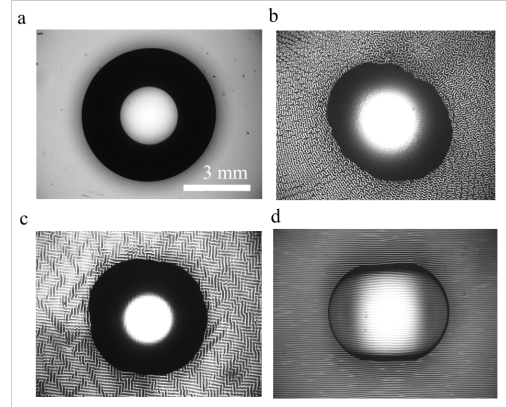


Figure 16: Optical microscopy images of droplets on (a) smooth, (b) fine-wrinkled, (c) Herringbone, and (d) striped surfaces.

These results collectively show that surface patterning can tune wetting anisotropy from negligible (smooth) to strong (striped), depending on the wrinkle geometry.

5 Conclusion

In this study, spectral methods solved the forward problem of the SHE, while an RBF-PINN addressed the inverse problem of parameter estimation. The project linked computational physics with real-world material applications by accurately characterising the parameters governing pattern formation.

The SHE was solved using Pseudo-Spectral methods to generate diverse pattern morphologies, enabling a 3D phase diagram that mapped physical parameters ε , δ , γ to their corresponding patterns. Furthermore, these patterns were produced as the input into the inverse system.

For the inverse problem, the RBF-PINN employed three-phase curriculum learning [12], GradNorm adaptive weighting [13] and the decomposition of the fourth-order PDE into two second-order PDEs. This approach mitigated large gradient losses, produced highly accurate image predictions, and successfully estimated the bifurcation parameter ε .

6 Outlook

The success of this framework showcases the RBF-PINN as a reliable tool for inverse parameter estimation in turing pattern-forming systems. However, there are several promising research directions that can be followed in future work. Inherently estimating all the parameters in the SHE ε , δ and γ would allow for full inverse identification of the underlying physics from a single pattern. This would require optimising and extending the PINN architecture for reliable multi-parameter discovery.

Additionally, mapping the SH parameters to measurable physical properties is another route that could be studied. This would allow the inverse design of wrinkled surfaces with desired functionalities such as hydrophobicity and anisotropic wetting.

Code & Data Availability

The code used in this paper is available for access on the following repositories:

<https://github.com/ZachMol/Wrinkling-Patterns-Project>
<https://github.com/ZachMol/SwiftHohenberg-Inverse-Pinn>
<https://github.com/ZachMol/Inverse-Pinn-Data-Analysis>

Acknowledgements

Thank you to all the support from Prof. J. Cabral, Dr. H. Lama, and the Polymer and Microfluids Group

References

- [1] Y. Peng, J. Shang, C. Liu, S. Zhao, C. Huang, Y. Bai, and Y. Li, “A universal replica molding strategy based on natural bio-templates for fabrication of robust superhydrophobic surfaces,” *Self-Collected Data (Source 1)*, 2023.
- [2] H. Lama, H. Mao, and J. T. Cabral, “Iso- to anisotropic wetting on microwrinkled surfaces with varying intrinsic wettability,” *Soft Matter*, vol. 21, p. 7125, 2025.
- [3] S. W. Cho and H. Son, “Physics-informed deep inverse operator networks for solving pde inverse problems,” *arXiv preprint arXiv:2412.03161*, 2024.
- [4] ProPERLA, “The lotus effect,” n.d. Accessed: 2025-12-08.
- [5] J.-T. Feng, F.-C. Wang, and Y.-P. Zhao, “Electrowetting on a lotus leaf,” *Biomicrofluidics*, vol. 3, no. 2, p. 022406, 2009.
- [6] skyblue16, “Water Drop On Rice Leaf.” iStock, 2018. Accessed: 2025-12-08.
- [7] Z. Huang, W. Hong, and Z. Suo, “Evolution of wrinkles in hard films on soft substrates,” *Physical Review E*, vol. 70, no. 3, p. 030601, 2004.
- [8] N. Stoop, R. Lagrange, D. Terwagne, P. M. Reis, and J. Dunkel, “Curvature-induced symmetry breaking determines elastic surface patterns,” *Nature Materials*, vol. 14, no. 3, pp. 337–342, 2015.
- [9] J. Burke and E. Knobloch, “Localized states in the generalized Swift-Hohenberg equation,” *Physical Review E*, vol. 73, no. 5, p. 056211, 2006.
- [10] A. El-Sayed, A. El-Mesady, and H. El-Saka, “A new fractional-order logistic equation with applications,” *Computers & Mathematics with Applications*, vol. 73, no. 6, pp. 1074–1082, 2017.
- [11] M. Cheng and J. Warren, “An efficient algorithm for solving the phase field crystal model,” *Journal of Computational Physics*, vol. 227, pp. 6241–6248, 2008.
- [12] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum learning,” *Proceedings of the 26th International Conference on Machine Learning (ICML)*, pp. 41–48, 2009.
- [13] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich, “Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks,” *arXiv preprint arXiv:1711.02257*, 2018.
- [14] A. Matas-Gil and R. G. Endres, “Unraveling biochemical spatial patterns: Machine learning approaches to the inverse problem of stationary turing patterns,” *iScience*, vol. 27, no. 6, p. 109822, 2024.
- [15] V. Shcherbakov and R. Schaback, “Scientific computing with radial basis functions,” *Frontiers in Numerical Analysis (Draft Chapter)*, 2009.
- [16] Institut für Theoretische Physik (Universität Münster), “Swift-Hohenberg Equation (Chapter 4).” [Lecture Script], 2010. Accessed: 25 October 2025.
- [17] D. S. Broomhead and D. Lowe, “Multivariable functional interpolation and adaptive networks,” *Complex Systems*, vol. 2, no. 3, pp. 321–355, 1988.
- [18] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, “Automatic differentiation in machine learning: A survey,” *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 5595–5637, 2018.
- [19] C. Isaac and K. Zareinia, “Effect of excessive neural network layers on overfitting,” *World Journal of Advanced Research and Reviews*, vol. 16, no. 02, pp. 1246–1257, 2022.

Predictive Viscosity Model for Sustainable Aviation Fuel

Eric Sihan Huang, Daanyaal Sajed

Department of Chemical Engineering, Imperial College London, U.K.

ABSTRACT

Sustainable aviation fuel is a promising alternative fuel, essential to decarbonising the aviation industry. A key bottleneck in its development is the lack of accurate predictive viscosity modelling. In this work, we address this gap by developing a fully predictive group-contribution viscosity model for linear and branched alkane mixtures. The proposed model is built upon Rosenfeld's residual entropy scaling framework and integrates the SAFT- γ Mie equation of state to accurately capture the relationship between the residual entropy and viscosity. A temperature-dependent extension of the residual entropy scaling approach is proposed to account for deviations from the ideal monovariate relationship observed for longer chain alkanes. To extend the model to mixtures, the Fang–He viscosity mixing rule is reformulated with a group-contribution approach. The proposed model is validated across a wide range of linear alkanes, branched alkanes, multi-component systems, and select aviation fuel blends. The viscosity predictions across all systems demonstrate high accuracy, with a percentage absolute average deviation of 2.87% for linear alkanes, 6.13% for branched alkanes and 5.18% for multi-component systems. Overall, the proposed methodology provides an accurate, predictive tool for viscosity prediction, optimised for sustainable aviation fuel blends.

1 INTRODUCTION

The aviation industry is a significant source of greenhouse gas emissions, accounting for approximately 2.5% of CO₂ emissions annually [1]. Thus, there is a clear need to decarbonise the aviation industry, and sustainable aviation fuel (SAF) represents a significant step towards achieving this goal. SAFs are projected to contribute to 65% of the reduction of emissions required by the aviation industry to reach net zero targets by 2050 [2]. These fuels are produced from sustainable feedstocks such as waste oils and fats, biomass, and synthetic products from carbon capture. Unlike petroleum-based jet fuels, where mining and refining fossil fuels emit CO₂, SAFs recycle carbon already absorbed in the biomass used in the feedstock [3].

Viscosity is a key fuel property with strict requirements. This is because viscosity governs the behaviour during fuel injection and atomisation, which greatly affects engine performance. Currently, the most widely used methods are empirical models, trained on experimental data [4]. However, gathering data for a large range of fluids over a range of conditions is time consuming. Therefore, development of more predictive models are essential in speeding up the transition to SAFs becoming more widely deployed [5].

Group-contribution (GC) models provide more versatility over empirical models. These models break down fluids into different functional groups, with each group contributing to the overall properties of the fluid. This enables models to predict properties of fluids without the need for experimental

data, by using previously trained group-specific parameters [6].

Aside from empirical fits, several theoretically based viscosity models have been developed. These include the free-volume theory [7] and the friction theory [8] among several others. While each of these models have their applications, a predictive liquid-viscosity model suitable for SAFs is still lacking. A large breakthrough in viscosity prediction was by Rosenfeld through his semi-empirical residual entropy scaling (RES) approach. Rosenfeld observed a monovariate relationship between a fluid's residual entropy and its transport properties, for simple fluids [9]. This RES method has been adopted and developed by several others, and is built upon in this work. While this monovariate collapse is highly accurate for simple, non-associating fluids, it lacks accuracy for longer and more complex molecules, such as those found in SAFs [10]. Thus in this work, we aim to adapt the RES approach for accurate SAF viscosity predictions.

In order to calculate the residual entropy, among other necessary thermodynamic properties, we integrate the Statistical Associating Fluid Theory (SAFT) equation of state (EoS). Specifically, SAFT- γ Mie, a state-of-the-art GC version of SAFT.

We focus on predicting the viscosities of branched and linear alkanes, since these are the groups that make up the majority of aviation blends, as displayed in Table 1.1. Specifically, the majority of alkanes in SAFs fall within the C9–C16 carbon number region [13], so while we aim to make our proposed model applicable to a wide range of alkanes, we place particular emphasis on this subset.

Table 1.1: Composition (wt.%) of various aviation fuels.

Composition (wt.%)	C-HRJ[11]	GTL[11]	Jet A[12]	Jet A-1[12]	HEFA-SPK[12]	ATJ-SPK[12]
n-Paraffin	14.2	34.8	28.1	19.6	11.7	-
Iso-Paraffin	63.1	56.3	38.8	29.9	87.3	99.8
Olefin	-	-	1.2	-	0.1	-
Naphthene	-	-	15.1	27.6	0.9	0.2
Aromatic	-	-	14.4	14.1	-	-
Other Compounds	22.7	8.9	2.4	8.7	-	-

Overall, we aim to develop a highly accurate predictive viscosity model for mixtures of branched and linear alkanes, by developing a GC approach and integrating the SAFT- γ Mie EoS.

The remainder of this report is structured as follows: in Section 2 we provide an overview of the theoretical background and methodology, detailing RES and the SAFT- γ Mie EoS, as well as previous models we adapt and build upon. In Section 3, we demonstrate the performance of our proposed model by validating against experimental data and with other predictive models. Finally, we summarise our findings and discuss possible future work.

2 METHODOLOGY

2.1 Rosenfeld's Residual Entropy Scaling

Residual Entropy Scaling (RES) is a semi-empirical approach which relates thermodynamic properties to transport properties. First postulated by Rosenfeld in 1977 [9], RES was developed from the observation of a monovariate relation between the residual entropy of a system and its transport properties, when expressed in a suitable dimensionless form. The residual entropy s_{res} is defined as the difference between the entropy of a fluid s and the ideal gas entropy s_{ideal} at a specific temperature T and molar density ρ ,

$$s_{\text{res}}(\rho, T) = s(\rho, T) - s_{\text{ideal}}(\rho, T). \quad (1)$$

The form of the dimensionless viscosity η^* proposed by Rosenfeld is,

$$\eta^* = \eta \frac{\rho_N^{-2/3}}{\sqrt{mk_B T}}, \quad (2)$$

where η is the viscosity, m is the mass of one molecule, k_B is the Boltzmann constant, ρ_N is the number density, and T is the temperature.

2.2 Advances in Reference Viscosity

Rosenfeld's scaling parameters perform well in the dense fluid region, however at low densities η^* diverges, preventing a monovariate collapse of the dimensionless viscosity to the residual entropy. To obtain a suitable scaling parameter for the entire fluid phase region, Novak proposed to reduce the

viscosity of a fluid by the dilute-gas viscosity $\eta_{\rho \rightarrow 0}$, commonly approximated by the Chapman-Enskog viscosity η_{CE} [14]. This reference produced a remarkably good collapse for n-alkanes and was adopted in the highly successful GC model of Lötgering-Lin *et al.* [15]. However, more recent theoretical analysis by Bell reveals that applying the dilute-gas limit for the reference viscosity does not yield the necessary reduction of units to conform with isomorph theory [16], a theory developed by Dyre which has elucidated the underlying physics behind RES [17].

Bell's solution to model the entire fluid phase was to reformulate the scaling by subtracting $\eta_{\rho \rightarrow 0}$ from η before applying Rosenfeld's dimensionless terms, essentially expressing the viscosity as the sum of an ideal viscosity and a residual viscosity,

$$\eta = \eta_{\rho \rightarrow 0} + \eta_{\text{res}}, \quad (3)$$

where η_{res} is then reduced by Rosenfeld's reference to yield the dimensionless residual viscosity, η_{res}^* . The core objective of this work is therefore to develop a GC method capable of predicting this quantity.

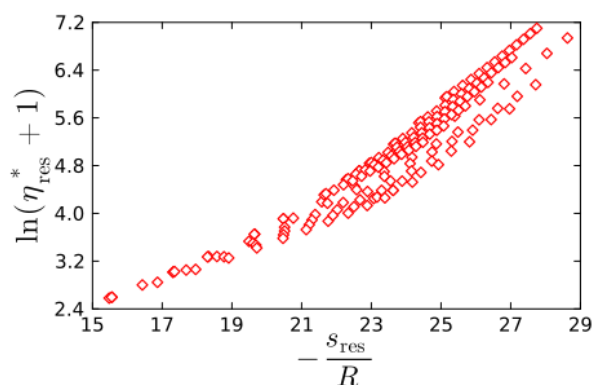


Figure 2.1: Dimensionless experimental data of squalane, with the dimensionless reference viscosity proposed by Bell plotted against dimensionless residual entropy, where R is the universal gas constant.

While Bell's reference is a substantial improvement to Rosenfeld's original reference, it does not fully address the issue of deviation from the monovariate function of residual entropy prevalent in constituent components of SAF. In particular, we find significant deviation in alkanes

with carbon lengths greater than decane (C10+), with a clear trend of further deviation with increasing chain lengths. For a given alkane, this experimental spread also increases with increasing residual entropy. This is illustrated in Figure 2.1.

Resolving this non-monovariate behaviour is key to accurately model the viscosity of SAFs, and requires the introduction of an additional parameter into the RES framework. We propose a temperature-dependent function τ to be introduced, of the form

$$\tau = 1 + \kappa \sqrt{T_r}, \quad (4)$$

with

$$T_r = \frac{T}{T_C}, \quad (5)$$

where T_r is the reduced temperature, T_C is the critical temperature, and κ is a fluid-specific coefficient which is implemented into our model as a GC variable. The effect of the function τ to successfully collapse η_{res}^* to a monovariate function is shown in Figure 2.2.

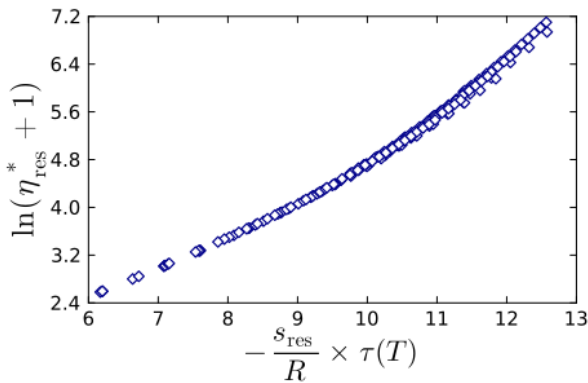


Figure 2.2: Dimensionless experimental data of squalane, utilising our proposed temperature-dependent function.

2.3 SAFT- γ Mie

In this work, a GC RES model is developed for SAF viscosity prediction. To ensure full predictive capability a GC EoS is employed, namely the SAFT- γ Mie EoS.

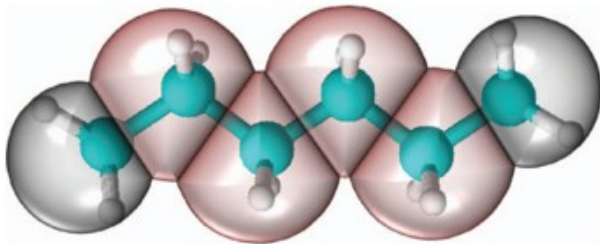


Figure 2.3: Figure from Papaioannou *et al.* showing pictorial representation of hexane as described within SAFT- γ Mie, with the grey indicating the CH₃ group and red the CH₂ group [18].

SAFT- γ Mie is an extension of the statistical associating fluid theory (SAFT) framework developed by Chapman *et al.* [19], that represents molecules as chains of functional groups with the interaction being described via the Mie potential, a generalised Lennard-Jones potential given as

$$\Phi_{kl}^{\text{Mie}}(r_{kl}) = C_{kl} \epsilon_{kl} \left[\left(\frac{\sigma_{kl}}{r_{kl}} \right)^{\lambda_{kl}^r} - \left(\frac{\sigma_{kl}}{r_{kl}} \right)^{\lambda_{kl}^a} \right], \quad (6)$$

with

$$C_{kl} = \frac{\lambda_{kl}^r}{\lambda_{kl}^r - \lambda_{kl}^a} \left(\frac{\lambda_{kl}^r}{\lambda_{kl}^a} \right)^{\frac{\lambda_{kl}^a}{\lambda_{kl}^r - \lambda_{kl}^a}}, \quad (7)$$

where ϵ_{kl} is the potential well depth, r_{kl} is the intersegment distance, λ_{kl}^r and λ_{kl}^a are the repulsive and attractive exponents, and σ_{kl} is the mean segment diameter. The coefficient C_{kl} ensures a minimum potential at $-\epsilon_{kl}$.

The SAFT- γ Mie EoS is expressed as the sum of contributions to the Helmholtz free energy A , the total A of a mixture takes the form

$$A = A_{\text{ideal}} + A_{\text{mono.}} + A_{\text{chain}} + A_{\text{assoc.}} \quad (8)$$

Here A_{ideal} is the free energy of an ideal gas mixture. $A_{\text{mono.}}$ is the free energy contribution accounting for Mie segment-segment interactions, calculated from a third-order Barker-Henderson perturbation expansion of the hard-sphere system [20]. A_{chain} is the free energy associated with the formation of fused monomers which make up the molecule, it is obtained through the application of Wertheim's Thermodynamic Perturbation Theory (TPT1) [21]. $A_{\text{assoc.}}$ accounts for the free energy contribution due to bonding site interactions, this is also calculated via Wertheim's TPT1.

For a complete mathematical description of the SAFT- γ Mie EoS, the reader is referred to Papaioannou *et al.* [22].

The RES approach relies on accurately determining s_{res} , this is achieved in the SAFT- γ Mie EoS through the following thermodynamic relation,

$$s_{\text{res}} = - \left(\frac{\partial A_{\text{res}}}{\partial T} \right)_{\rho}, \quad (9)$$

with

$$A_{\text{res}} \equiv A - A_{\text{ideal}} = A_{\text{mono.}} + A_{\text{chain}} + A_{\text{assoc.}} \quad (10)$$

In this work, thermodynamic calculations are carried out through the open-source thermodynamic package Clapeyron.jl [23].

2.4 Entropy Scaling Model

This work proposes a GC viscosity model geared towards the branched and linear alkanes which

constitute the majority of SAFs. The proposed model is developed from the RES framework, where we incorporate a temperature dependence function, as discussed in Section 2.2.

The empirical relation to determine η_{res}^* is given by a third-order polynomial expansion,

$$\ln(\eta_{\text{res}}^* + 1) = A + Bz + Cz^2 + Dz^3, \quad (11)$$

with

$$z = -\frac{s_{\text{res}}}{Rm_{\text{gc},S}}, \quad (12)$$

where $m_{\text{gc},S}$ is the sum of the shape factors S_k of the constituent groups. The division by $m_{\text{gc},S}$ is adapted from the work of Lötgering-Lin *et al.*, and acts as a normalisation term to ensure that molecules of various molecular sizes have similar values of the dimensionless residual entropy.

The coefficients of the polynomial A , B , C , and D are expressed in terms of group-specific parameters A_i , B_i , C_i , and D_i as

$$A = \sum_{i=1}^{N_G} n_i S_i \sigma_i^3 A_i, \quad (13)$$

$$B = \sum_{i=1}^{N_G} \frac{n_i S_i \sigma_i^3}{V_{\text{tot}}^\gamma} B_i, \quad (14)$$

$$C = \sum_{i=1}^{N_G} n_i C_i \tau, \quad (15)$$

$$D = \sum_{i=1}^{N_G} n_i S_i \sigma_i^3 D_i, \quad (16)$$

with

$$V_{\text{tot}} = \sum_{i=1}^{N_G} n_i S_i \sigma_i^3, \quad (17)$$

$$\tau = 1 + \kappa \sqrt{T_r}, \quad (18)$$

$$\kappa = \kappa_0 + \kappa_1 m_{\text{gc},S}, \quad (19)$$

where τ is the factor introduced in Section 2.2, n_i is the number of groups i in the fluid, N_G is the total number of groups in the fluid, and γ is a global constant. The coefficient κ is correlated as a linear function of $m_{\text{gc},S}$, where κ_0 and κ_1 are global constants.

Note that the temperature-dependent function τ is introduced into the quadratic coefficient C rather than being applied to the entire reference η_{res}^* . This approach is adopted for two main reasons. Firstly, we find that altering η_{res}^* would necessitate a new, more complex, GC mixing rule framework. Secondly, τ is introduced to correct the non-monovariate behaviour, and placing this correction term on a higher order term is consistent with our observations of increasing deviation with residual entropy.

Furthermore, the cubic term D is expanded to be a GC variable. As opposed to the global constant

approach of Lötgering-Lin *et al.*, which provides the necessary flexibility to model the wide range of target alkanes in SAFs.

2.5 Liquid Mixtures

In this work we build upon the Fang-He viscosity mixing rule [24]:

$$\ln(\eta_{\text{mix}}) = \sum_{c=1}^{N_C} \theta_c \ln(\eta_c) + \sum_{c=1}^{N_C} x_c \ln\left(\frac{\theta_c}{x_c}\right) + \left(\sum_{c=1}^{N_C} x_c v_c^{2/3}\right) \sum_{j < k} \frac{w_{j,k}}{RT} \theta_j \theta_k, \quad (20)$$

where η_{mix} is the mixture viscosity, $w_{j,k}$ is the binary exchange parameter between components j and k , N_C is the total number of components in the mixture, θ_c is the molecular surface fraction, and v_c is the molar volume for given component c .

Fang *et al.* developed this from Arrhenius' viscosity mixing rule [25], and adapted it for non-ideal mixtures based upon Eyring's absolute reaction rate theory, with a modified version of the Flory-Huggins equation [24].

The molecular surface fraction can be calculated from quantum chemistry calculations. However, this would complicate the model, and Fang *et al.* sought to develop an approach that remained accurate yet simple. They therefore adopted a hard-sphere model, which approximates the surface area of a molecule, with its molar volume v_c . The molecular surface fraction of a component in a mixture is therefore given as,

$$\theta_c = \frac{x_c v_c^{2/3}}{\sum_{c=1}^{N_C} x_c v_c^{2/3}}. \quad (21)$$

Fang *et al.* calculated the binary exchange parameters, $w_{j,k}$, by individually optimising for each binary system from experimental data [24]. We adapt this to a GC approach. First, a component-specific parameter w_c is calculated for each fluid. This is done by a summation in terms of the number of groups n_i and its group-specific parameter $w_{G,i}$, summing across all groups in the compound,

$$w_c = \sum_{i=1}^{N_G} w_{G,i} n_i. \quad (22)$$

Then the binary exchange parameter $w_{j,k}$ for each binary pair is found by a weighted sum of each component's w_c and θ_c ,

$$w_{j,k} = \theta_j w_j + \theta_k w_k. \quad (23)$$

2.6 Optimisation of Model Parameters

The parameters in our model are obtained via an optimisation of the objective function:

$$\min_{A,B,C,D,\gamma,\kappa_0,\kappa_1} \frac{1}{N} \sum_{i=1}^N \left(\frac{\eta_{i,\text{exp}} - \eta_{i,\text{calc}}}{\eta_{i,\text{exp}}} \right)^2, \quad (24)$$

where N is the number of data points, $\eta_{i,\text{exp}}$ is an experimental datum and $\eta_{i,\text{calc}}$ is the predicted viscosity.

To perform this complex, high-order optimisation, the Covariance Matrix Adaption Evolutionary Strategy (CMA-ES) optimisation algorithm was employed [26].

To target the model towards SAFs, each fluid has a different weighting within the optimisation. This weighting was chosen based on several factors; such as relevance to SAFs, number of data points, number of data sources, and experimental uncertainty. A comprehensive list of the data sets and their associated weightings is provided in the supplementary information.

3 RESULTS AND DISCUSSION

In this section we present the results of our proposed model, using the optimised GC parameters in Table 3.1 and the global parameters in Table 3.2. The performance is compared to other GC models, namely, the models proposed by Lötgering *et al.* [15] and Mickoleit *et al.* [27] in Figures 3.1 and 3.3. The Mickoleit *et al.* model is a GC method which also implements the reference viscosity developed by Bell and utilises the PC-SAFT EoS, but only has one parameter per group. Note that not all models are shown for all fluids, as both Lötgering *et al.* and Mickoleit *et al.* report results of their respective models on different fluids.

The accuracy of the models are quantified by calculating the percentage absolute average deviation %AAD between the experimental data and predicted viscosities,

$$\%AAD = \frac{100}{N} \sum_{i=1}^N \frac{|\eta_{i,\text{exp}} - \eta_{i,\text{calc}}|}{\eta_{i,\text{exp}}}. \quad (25)$$

The reported experimental data uncertainty varies between datasets, but the most accurate viscosity measurement techniques can achieve uncertainties of 2.0% [28]. Unfortunately, some systems exhibit larger uncertainties. The most extreme case of this is 2,2,4 – trimethylhexane, which reports uncertainties of up to 9.0% [29]. We minimise the use of less reliable data wherever possible; however, it may still contribute to certain compounds performing worse than expected.

Table 3.1: Proposed model group-contribution parameters A_i , B_i , C_i , and D_i for each group.

Group	$A_i \times 10^3$	$B_i \times 10^3$	$C_i \times 10^2$	$D_i \times 10^3$
CH ₃	-2.1974	1.2106	-6.8615	0.1713
CH ₂	-2.4598	0.3496	-0.9086	0.1590
CH	19.2750	-4.0447	6.0865	1.6680
C	37.4715	-10.4740	13.8382	3.4487

Table 3.2: Proposed model global parameters γ , κ_0 and κ_1 .

Global Parameter	Value
γ	-0.3587
κ_0	-1.3403
κ_1	0.5041

3.1 Viscosity of Linear Alkanes

Table 3.3: Summary of linear alkane datasets and calculated viscosity %AAD using our proposed model.

Substance	Number of Data Points	%AAD
n-butane	347	3.01
n-pentane	600	1.74
n-hexane	367	2.26
n-heptane	290	4.43
n-octane	453	3.03
n-nonane	289	2.79
n-decane	358	2.44
n-undecane	231	0.91
n-dodecane	354	1.81
n-tridecane	250	2.91
n-tetradecane	167	3.69
n-pentadecane	204	2.95
n-hexadecane	241	4.25
n-heptadecane	582	2.65
n-octadecane	121	2.29
n-nonadecane	111	2.83
n-eicosane	61	3.06
Mean		2.87

The predictive capability of our proposed model is summarised in Table 3.3, which presents the %AAD of linear alkanes across the investigated range. This range fully encompasses the linear alkanes relevant to SAFs. The model demonstrates a high level of accuracy with a mean %AAD of 2.87%.

A comparison of our model's performance to other GC models is presented in Figure 3.1. The model slightly improves upon the already highly accurate model of Lötgering *et al.*, where a mean %AAD of 3.67% is reported for this range of linear

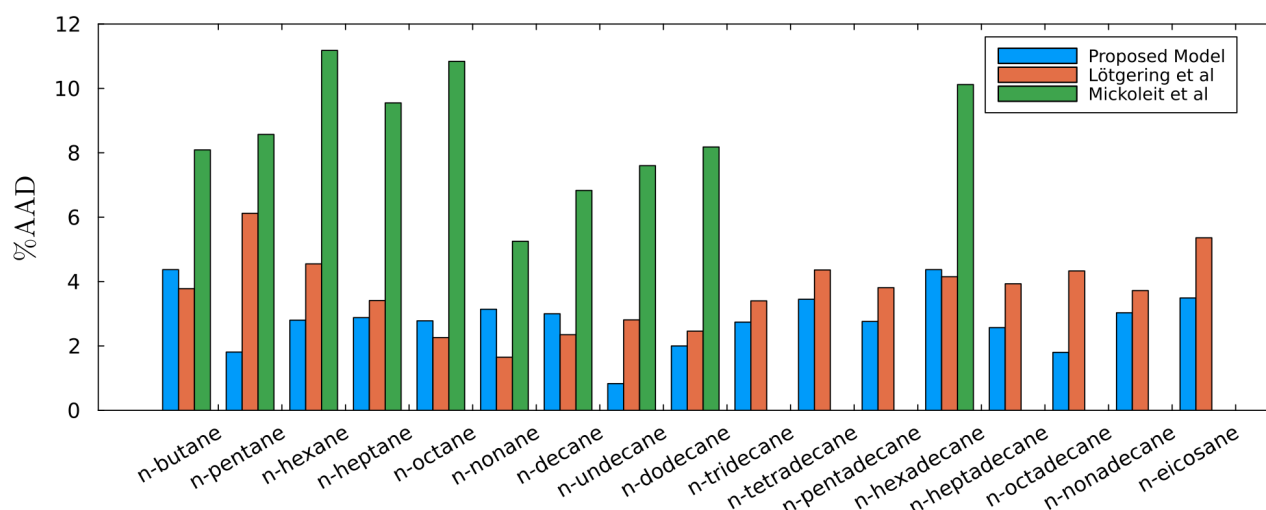


Figure 3.1: Errors (%AAD) of predicted viscosities of linear alkanes using our proposed model compared with Lötgering *et al.* [15] and Mickoleit *et al.* [27] models.

alkanes. Furthermore, the model's predictive ability to capture the viscosity over a range of temperatures is demonstrated for a select set of linear alkanes in Figure 3.2.

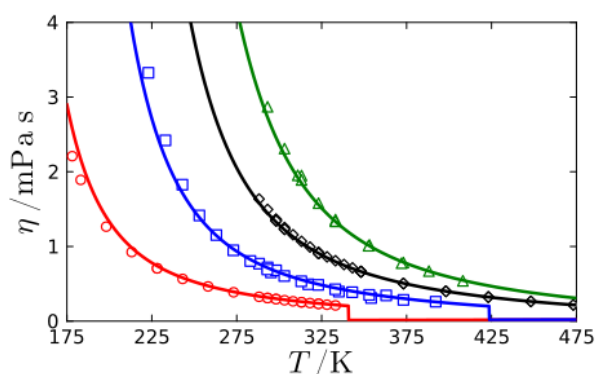


Figure 3.2: Atmospheric viscosity-temperature of selected linear alkanes, $\circ$ -hexane, $\square$ -nonane, $\diamond$ -dodecane, $\triangle$ -pentadecane.

3.2 Viscosity of Branched Alkanes

Table 3.4 summarises our model's ability to predict the viscosity of branched alkanes, with a comparison to previous models in Figure 3.3. A major focus of this work is to improve the predictive capability of branched alkanes, as they have performed notably poor in previous GC models. As discussed prior, this arises from the assumption of a monovariate relationship between residual entropy and dimensionless viscosity failing for highly branched fluids.

Our results show improved performance for a wide range of branched alkanes, with an overall mean %AAD of 6.13%. We find substantial improvement over the other models in longer alkanes, particularly for squalane, over the Lötgering *et al.* model, but

Table 3.4: Summary of pure-component branched alkane datasets and calculated viscosity %AAD using our proposed model.

Substance	Number of Data Points	%AAD
2,6,10,14-tetramethylpentadecane	30	15.21
2-methylpropane	139	12.67
2-methylbutane	154	3.22
2-methylpentane	49	3.32
3-methylpentane	78	2.00
squalane	170	4.68
3,5-dimethylheptane	9	6.24
2-methyloctane	10	2.12
4-methyloctane	10	9.47
3,6-dimethyloctane	10	1.85
2-methylnonane	10	1.69
2-methyldecane	10	0.99
3-methylundecane	10	1.28
2-methylpentadecane	10	0.74
7-methylhexadecane	10	7.90
2,2,4-trimethylpentane	345	5.77
isocetane	90	6.49
2,2,4-trimethylhexane	13	12.81
2,4-dimethylpentane	6	9.21
3-ethylpentane	6	14.94
Mean		6.13

also for relatively short branched alkanes such as 2-methylpentane. However, a select few fluids have higher errors, a factor contributing to this is the lack of accurate experimental data as the poorer performing fluids have limited experimental data points, with many from only one source. While the %AAD results are slightly worse than that of linear alkanes, Figure 3.4 demonstrates that the model

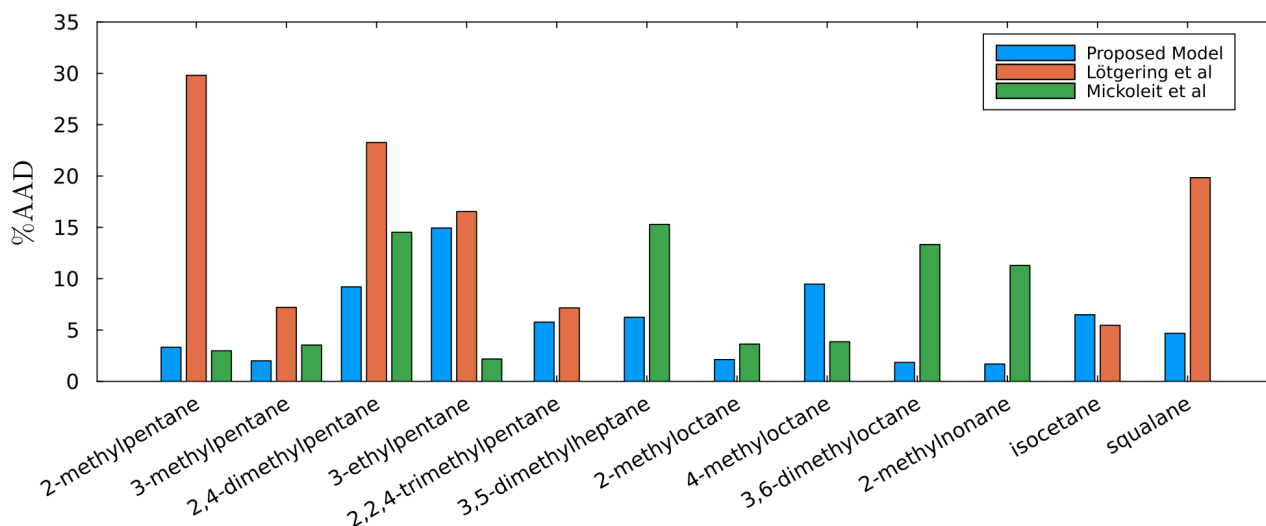


Figure 3.3: Errors (%AAD) of predicted viscosities of branched alkanes using our proposed model compared with Lötgering *et al.* [15] and Mickoleit *et al.* [27] models.

remains highly accurate for a range of branched alkanes at atmospheric pressure.

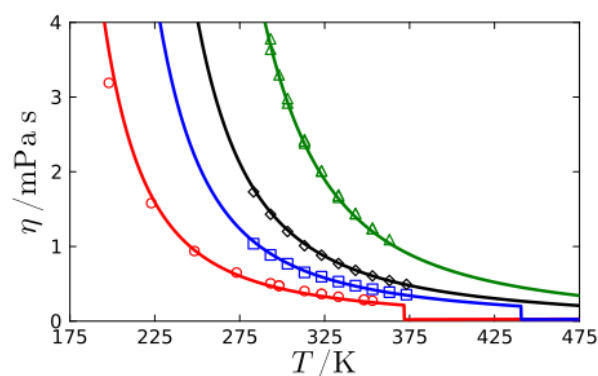


Figure 3.4: Atmospheric viscosity-temperature of selected branched alkanes, $\circ$ -2,2,4-trimethylpentane, $\square$ -2-methylnonane, $\diamond$ -3-methylundecane, $\triangle$ -isocetane.

3.3 Viscosity of Mixtures

Table 3.5: Mixture group-contribution parameters w_i for each group.

Group	$w_{G,i} \times 10^{-3}$
CH ₃	-99.590
CH ₂	15.123
CH	35.796
C	187.149

The proposed pure-component viscosity model is extended to mixtures by incorporating our GC adaption of the Fang-He mixing rule, using the parameters in Table 3.5. A summary of the results, in terms of the %AAD, are presented in Table 3.6.

Table 3.6: Summary of mixture datasets and calculated viscosity %AAD using our proposed model.

Mixture	Number of Data Points	%AAD
isocetane + n-dodecane	90	1.47
2,2,4-trimethylpentane + n-dodecane	30	9.92
2,2,4-trimethylpentane + n-octane	32	9.47
n-dodecane + n-heptadecane	24	3.28
n-eicosane + n-heptane	32	3.47
n-hexane + n-dodecane	292	3.70
n-pentane + n-octane	279	4.74
squalane + n-octane	80	8.36
squalane + n-hexane	96	9.27
n-tridecane + isocetane	252	2.28
n-tridecane + n-pentadecane	22	3.76
2,2,4-trimethylpentane + n-heptane	88	3.92
n-butane + n-heptane	7	5.04
n-decane + n-heptane	311	5.58
n-decane + n-hexadecane	54	2.34
n-heptane + n-nonane	57	2.66
n-hexadecane + n-hexane	174	10.33
n-decane + n-eicosane + n-tetracosane	18	4.97
n-pentane + n-octane + n-decane	530	6.40
n-decane + n-dodecane + n-tetradecane + n-hexadecane	90	2.72
Mean		5.18

Our proposed model achieves high accuracy at predicting viscosities over a wide range of mixtures of linear and branched alkanes, with a mean %AAD of 5.18%.

While binary mixture data are readily available, reported viscosity data for higher-order systems of alkanes are limited. Consequently, we can only validate our model with a small number of higher-order systems. Regardless, as shown in Table 3.6, these mixtures still perform well, indicating that the model extends to other higher-order systems.

Figures 3.5 and 3.6 display the predicted viscosities of two different mixtures at various conditions, along with their respective experimental data. These show that our model holds well at a range of temperatures, pressures and compositions, for mixtures of linear and branched alkanes.

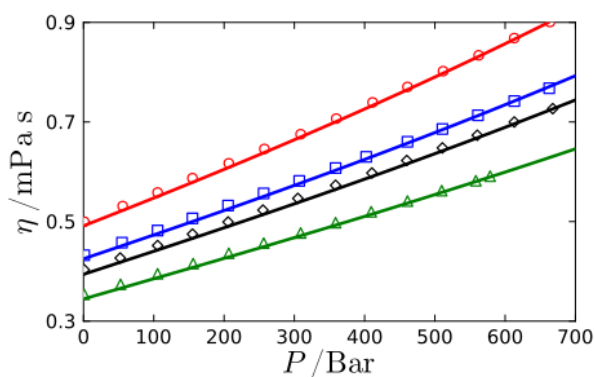


Figure 3.5: Predicted viscosities and experimental data of binary heptane + hexane mixtures at varying temperatures and compositions, plotted against pressure.
 ○-T=303K, $x_{\text{heptane}}=0.46$, □-T=303K, $x_{\text{heptane}}=0.75$,
 ◇-T=323K, $x_{\text{heptane}}=0.46$, △-T=323K, $x_{\text{heptane}}=0.75$.

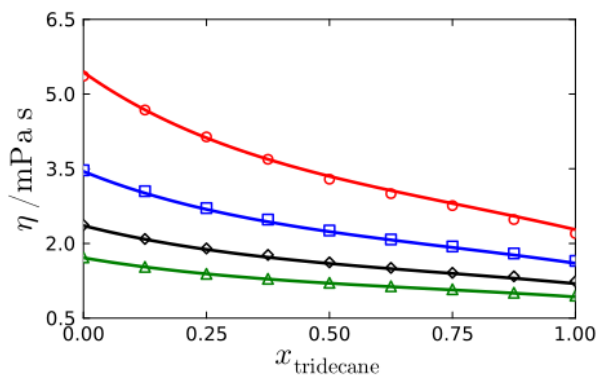


Figure 3.6: Predicted viscosities and experimental data of binary tridecane + isocetane mixtures at varying temperatures, plotted against composition.
 ○-T=293.15K, □-T=313.15K, ◇-T=333.15K, △-T=353.15K.

3.3.1 Viscosity of Aviation Fuel Mixtures

To ensure our model can accurately predict the viscosities of aviation fuel mixtures we compare it to experimental data for two fuel blends. The reported mixtures are a blend of 18 different linear and branched alkanes for GTL and 16 components for C-HRJ, while the remainder of each was reported as cycloparaffins and naphthalenes [11]. As the

specific components are not available we normalise the reported composition.

Figure 3.7 and Table 3.7 show that for GTL our model performs very well, but is slightly worse for C-HRJ. This error is likely due to the normalisation of the fuel composition, as only 77.29% of C-HRJ is known, while 91.10% for GTL. The model's under-prediction is expected, given that the unreported components in aviation fuel are heavier fluids with high viscosities [30, 31].

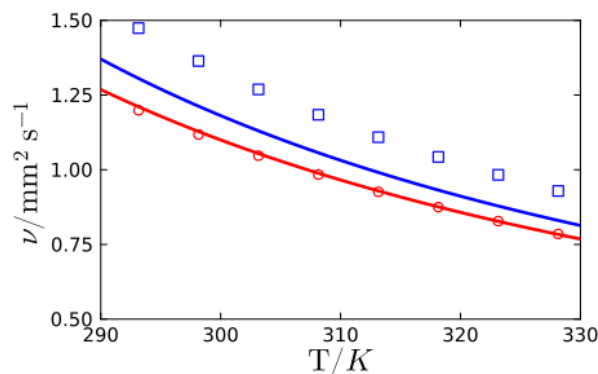


Figure 3.7: Predicted kinematic viscosities and experimental data of aviation fuel mixtures ○-GTL and □-C-HRJ.

Table 3.7: %AAD for mixture viscosity predictions.

Mixture	%AAD
GTL	1.0
C-HRJ	10.8

3.4 Performance of Proposed GC Methods

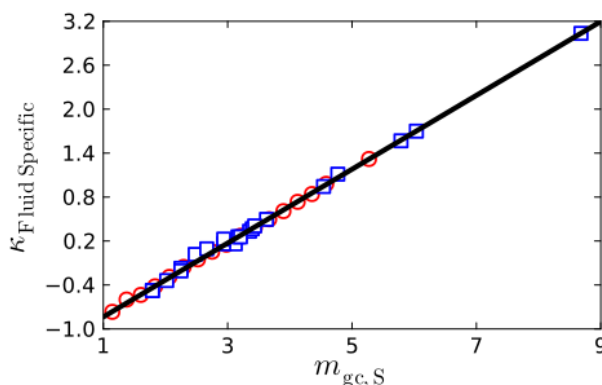


Figure 3.8: Individually optimised fluid specific κ parameter (y-axis) plotted against $m_{gc,S}$ (x-axis), where ○-T=Linear Alkanes and □-Branched Alkanes.

To compare the performance of a GC-based and mixture-specific κ in the proposed temperature dependency (Equations 18 and 19) we plot the individually optimised κ parameter for each fluid

against the fluid's $m_{gc,S}$. As shown in Figure 3.8, it is clear there is a very strong linear relation. This suggests that our proposed linear GC method is highly accurate as it can successfully predict the correct κ for the wide range of both linear and branched alkanes tested.

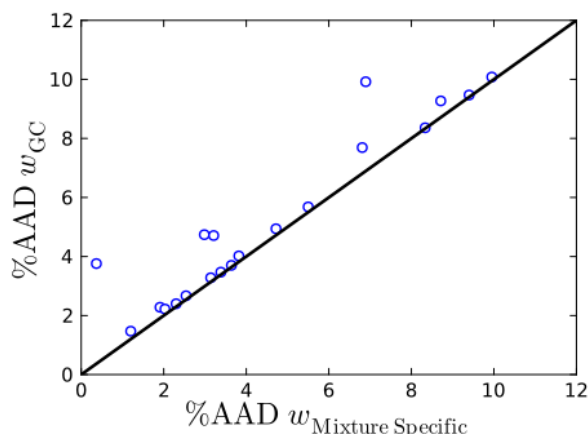


Figure 3.9: Comparison of errors (%AAD) for predicting viscosities of mixtures when using a group-contribution w parameter (y-axis) against errors using individually trained, mixture specific w parameters (x-axis).

Similarly, to test the effectiveness of our proposed GC adaption of the Fang-He mixing rule we compare the %AAD from a mixture specific w parameter and our GC method, as displayed in Figure 3.9. This shows that our GC approach can accurately calculate the optimum $w_{j,k}$ parameter for the majority of mixtures tested, with only a few outliers where the approach loses accuracy. However, even in these cases the GC approach does not exhibit a notably larger error.

4 CONCLUSION

In this study, we demonstrate that by incorporating a temperature dependence, a monovariate relation between residual entropy and viscosity can be formed for a wide range of fluids. From this, applying a third-order polynomial expansion with the proposed GC parameters enables accurate liquid viscosity predictions. Furthermore, we show that incorporating a GC parameter into the Fang-He mixing rule can similarly accurately predict the viscosities of mixtures of branched and linear alkanes. This approach requires five GC variables per group, but crucially does not require component specific parameters, thus is entirely predictive, requiring no experimental data for new fluids.

5 FUTURE WORK

As we have trained our model on branched and linear alkanes, the next logical step would be to expand the model to include groups found in other

components in SAF, such as aromatic and cyclic groups.

Due to our use of SAFT- γ Mie, the model does not differentiate between certain isomers if they contain the same number of each group. Thus it lowers the potential accuracy for branched alkanes. This could be corrected by utilising a version of SAFT which can differentiate between these types of isomers, such as structural (s)-SAFT- γ Mie [32].

6 SUPPORTING INFORMATION

Access to raw experimental data, optimisation codes, and example calculations are available at: <https://github.com/erichuang384/Supporting-Information-Huang-Sajed-2025>

REFERENCES

- [1] H. Ritchie. What share of global CO emissions come from aviation? *Our World in Data* (2024). <https://archive.ourworldindata.org/20251125-173858/global-aviation-emissions.html>.
- [2] International Air Transport Association (IATA). *Policy: Net Zero CO Emissions Roadmap*. 2024. <https://www.iata.org/contentassets/8d19e716636a47c184e7221c77563c93/policy-net-zero-roadmap.pdf> (visited on 12/05/2025).
- [3] Environmental & Energy Study Institute (EESI). *An Introduction to Sustainable Aviation Fuels*. 2023. https://www.eesi.org/articles/view/an-introduction-to-sustainable-aviation-fuels?utm_source=chatgpt.com (visited on 12/10/2025).
- [4] H. O. Baled et al. Viscosity models for pure hydrocarbons at extreme conditions: A review and comparative study. *Fuel* **218** (2018), pp. 89–111. <https://www.sciencedirect.com/science/article/pii/S0016236118300024>.
- [5] M. Gülüm. Modelling the Viscosity-Temperature Relationship of Alternative Fuel Blends: Comparison of Empirical and Machine Learning Models. *International Journal of Thermophysics* **46**(6) (2025), p. 85. ISSN: 1572-9567. <https://doi.org/10.1007/s10765-025-03556-y>.
- [6] B. P. Agbodekhe et al. Enhanced thermophysical property prediction with uncertainty using group-contribution models. *Molecular Systems Design & Engineering* (2025). <https://doi.org/10.1039/d5me00126a>. <https://pubs.rsc.org/en/content/articlehtml/2025/me/d5me00126a>.
- [7] F. Llovel, R. M. Marcos & L. F. Vega. Free-Volume Theory Coupled with Soft-SAFT for Viscosity Calculations: Comparison with Molecular Simulation and Experimental Data. *J. Phys. Chem. B* **117**(27) (July 2013), pp. 8159–8171. <https://doi.org/10.1021/jp401307t>.
- [8] S. E. Quiñones-Cisneros, C. K. Zéberg-Mikkelsen & E. H. Stenby. The friction theory (f-theory) for viscosity modeling. *Fluid Phase Equilibria* **169**(2) (2000), pp. 249–276. <https://www.sciencedirect.com/science/article/pii/S0378381200003101>.

- [9] Y. Rosenfeld. Relation between the transport coefficients and the internal entropy of simple systems. *Phys. Rev. A* **15**(6) (June 1977), pp. 2545–2549. <https://link.aps.org/doi/10.1103/PhysRevA.15.2545>.
- [10] Ian H. Bell. Probing the link between residual entropy and viscosity of molecular fluids and model potentials. *Proceedings of the National Academy of Sciences* **116**(8) (2019), pp. 3373–3380. <https://doi.org/10.1073/pnas.1816512116>.
- [11] T. J. Fortin & A. Laesecke. Viscosity Measurements of Aviation Turbine Fuels. *Energy & Fuels* **29**(9) (2015), pp. 5495–5506. <https://doi.org/10.1021/acs.energyfuels.5b00423>.
- [12] M. F. Rojas-Michaga et al. “Techno-economic and life cycle assessment review of sustainable aviation fuel produced via biomass gasification”. *Sustainable Biofuels: Opportunities and Challenges*. Ed. by R. C. Ray. 2021, pp. 269–303. <https://doi.org/10.1016/B978-0-12-820297-5.00012-8>.
- [13] M. A. Díaz-Pérez & J. C. Serrano-Ruiz. Catalytic Production of Jet Fuels from Biomass. *Molecules* **25**(4) (2020), p. 802. <https://doi.org/10.3390/molecules25040802>.
- [14] L. T. Novak. Fluid Viscosity–Residual Entropy Correlation. *International Journal of Chemical Reactor Engineering* **9** (2011), A107.
- [15] O. Lötgering-Lin & J. Gross. Group Contribution Method for Viscosities Based on Entropy Scaling Using the Perturbed-Chain Polar Statistical Associating Fluid Theory. *Industrial & Engineering Chemistry Research* **54**(32) (2015), pp. 7942–7952. <https://doi.org/10.1021/acs.iecr.5b01698>.
- [16] I. H. Bell. Entropy Scaling of Viscosity—II: Predictive Scheme for Normal Alkanes. *Journal of Chemical & Engineering Data* **65**(11) (2020), pp. 5606–5616. <https://doi.org/10.1021/acs.jced.0c00749>.
- [17] J. C. Dyre. Perspective: Excess-entropy scaling. *The Journal of Chemical Physics* **149**(21) (2018), p. 210901. <https://doi.org/10.1063/1.5055931>.
- [18] V. Papaioannou et al. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments. *The Journal of Chemical Physics* **140**(5) (2014), p. 054107. <https://doi.org/10.1063/1.4851455>.
- [19] W. G. Chapman, G. Jackson & K. E. Gubbins. Phase equilibria of associating fluids. *Molecular Physics* **65** (1988), pp. 1057–1079.
- [20] A. J. Haslam et al. Expanding the Applications of the SAFT- Mie Group-Contribution Equation of State: Prediction of Thermodynamic Properties and Phase Behavior of Mixtures. *Journal of Chemical & Engineering Data* **65**(12) (2020), pp. 5862–5890. <https://doi.org/10.1021/acs.jced.0c00746>.
- [21] M. S. Wertheim. Thermodynamic perturbation theory of polymerization. *The Journal of Chemical Physics* **87** (1987), pp. 7323–7331.
- [22] V. Papaioannou et al. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments. *The Journal of Chemical Physics* **140**(5) (2014), p. 054107. <https://doi.org/10.1063/1.4851455>.
- [23] P. J. Walker, H.-W. Yew & A. Riedemann. Clapeyron.jl: An Extensible, Open-Source Fluid Thermodynamics Toolkit. *Ind. Eng. Chem. Res.* **61**(20) (2022), pp. 7130–7153. <https://pubs.acs.org/doi/10.1021/acs.iecr.2c00326>.
- [24] S. Fang. Viscosity Model for Mixtures Based on Eyring's Absolute Reaction Theory. *Progress in Chemistry* **22**(0203) (2010), pp. 309–314. https://www.researchgate.net/publication/286148518_Viscosity_Model_for_Mixtures_Based_on_Eyring's_Absolute_Reaction_Theory.
- [25] N. Gao et al. Viscosity of Ionic Liquids: Theories and Models. *Chemical Reviews* **124**(1) (2024), pp. 27–123. <https://doi.org/10.1021/acs.chemrev.3c00339>.
- [26] N. Hansen. The CMA Evolution Strategy: A Tutorial. *arXiv preprint arXiv:1604.00772* (2016). arXiv:1604.00772 [cs.LG].
- [27] E. Mickoleit et al. Group Contribution Method for the Residual Entropy Scaling Model for Viscosities of Branched Alkanes. *International Journal of Thermophysics* **44**(12) (2023), p. 176. <https://doi.org/10.1007/s10765-023-03289-w>.
- [28] M. Trusler & I. Eiuubovi. Investigating Viscosity and Density of Hydrocarbon Liquids with Dissolved Gases through Experiments and Modelling for CO₂ Sequestration in Depleted Reservoirs. *Proceedings of the 2025 AIChE Annual Meeting 2025*.(721530) (2025), -. <https://aiche.confex.com/aiche/2025/meetingapp.cgi/Paper/721530>.
- [29] S. Bair. The pressure dependence of viscosity for 2,2,4 trimethylhexane to 1 GPa along the 20 °C isotherm. *Fluid Phase Equilibria* **488** (2019), pp. 9–12. <https://www.sciencedirect.com/science/article/pii/S0378381219300299>.
- [30] Thomas J. Bruno. *Composition-Explicit Distillation Curves for Aviation Fuel Blends*. Tech. rep. NIST Technical Note 1664. National Institute of Standards and Technology, 2010. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=904586.
- [31] Thomas J. Bruno & Evgenii Baibourine. Comparison of Biomass-Derived Turbine Fuels with the Composition-Explicit Distillation Curve Method. *Energy & Fuels* **25**(4) (2011), pp. 1847–1858. <https://doi.org/10.1021/ef200115b>. <https://doi.org/10.1021/ef200115b>.
- [32] Arne Schulze-Hulbe et al. Extending the Structural (s)-SAFT- γ Mie Equation of State to Primary Alcohols. *Industrial & Engineering Chemistry Research* **61**(33) (2022), pp. 12208–12228. <https://doi.org/10.1021/acs.iecr.2c00198>.

Optimising Hydrogen Deployment for Long-Duration Energy Storage in Future UK Electricity Networks

Michael Joshua and Shaylan Jasani

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

The UK's transition to a renewable-dominated electricity network requires consideration of how increased generator capacities interact with seasonal storage to manage periods of energy deficit. This study investigates the current and future prospects of long-duration energy storage (LDES) potential of green hydrogen for decarbonisation of the UK electricity network by 2050. Python for Power System Analysis (PyPSA) was used to develop models including thermal, renewable, and nuclear generation technologies, together with lithium-ion batteries and underground hydrogen storage. Greenfield optimisation models were created to assess the viability of hydrogen over single and multiple years, addressing differences due to inter-annual variability. Myopic pathway optimisation models were then developed for single and multi-year periods, accounting for variable costs and unit lifetimes. This provided a decennial development pathway for the UK electricity network from 2020 until 2050. Results showed that hydrogen deployment occurs in later decades, with non-pathway models exhibiting introduction in 2040. By 2050, the maximum energy capacity requirements for single and multi-year hydrogen storage reached 9.93 TWh and 11.5 TWh, respectively, which are 2.87% and 3.33% of the total average annual energy consumption. In contrast, the pathway models delayed hydrogen usage until 2050 and required a significantly lower capacity (2.51 TWh) for the single-year case, but a comparable 11.0 TWh for the multi-year model. Underground hydrogen storage proved more cost-effective than tank storage, as it displayed greater utilisation under identical conditions. The resulting generation mixes are renewable-dominated by 2050, with offshore wind providing the greatest contribution.

1 Introduction

The UK has made commitments to achieve net zero greenhouse gas (GHG) emissions by 2050, and the strategies implemented to achieve this place a heavy emphasis on renewable energy sources. Reworking electricity systems to be dominated by wind and solar PV generation has proven to be an economically viable method to decarbonise regions local to the site of generation. As costs of building renewable generators have been decreasing continually during the last decade, these sources are now cost-competitive with traditional non-renewable resources including oil and gas. The now cheaper sources, particularly onshore wind and solar, have shown 75% and 90% respective increases in deployment from 2015 to 2024. Offshore wind is a promising technology for the future of the UK electricity network, displaying greater reliability than onshore wind and solar. The initial capital investment is currently higher than the other sources, however it will become more cost-competitive approaching 2050 [1].

National-scale electricity systems that are dominated by renewable generation must account for daily and seasonal variability in weather conditions to be reliable in meeting demand. The role of energy storage within a renewable-dependent electricity grid is to mitigate these variations by storing excess energy during a surplus to be utilised during

a deficit of renewable generation. Currently, batteries are at the focus of energy storage projects as they are relatively easy to deploy and can effectively accommodate short-term variations. However, batteries are not suitable for mitigating long-term, seasonal variation in generation due to significant costs of retaining charge over a sustained period of time [2]. Therefore, LDES is necessary to accommodate for the seasonal changes of demand with renewable generation.

Currently, the UK has an installed capacity of 26.7 GWh of pumped-hydro storage (PHS) [3], which accounts for nearly all of the LDES capacity. Thermal energy storage (TES) and green hydrogen storage are emerging technologies that could supplement PHS in periods of extreme reduced generation. Seasonal hydrogen LDES uses renewable energy to produce green hydrogen through the electrolysis of water, which is subsequently stored. This hydrogen can then be converted back to electricity using turbines for use during a deficit. The electrolysis of water has an efficiency of 80% and the hydrogen turbines have efficiencies of 50%, leading to a total round trip efficiency (RTE) of 40% [4]. The low RTE, coupled with high capital investment for both electrolyzers and hydrogen turbines, can currently act as a barrier to entry for hydrogen LDES.

Due to the inter-annual variability of weather and hence re-

newable generation, it is important to use multi-year series to more accurately model this and the associated uncertainty with rare and extreme weather events. This is supported by existing literature [5], and is crucial to develop a robust renewable energy system for the UK and has a direct impact on the utilisation of LDES.

This study aims to provide an insight into the future long duration energy storage requirements for a renewable-dominant UK grid which achieves the net zero emissions targets via the holistic transition pathway outlined by the National Energy System Operator (NESO) [6]. The models in this study were developed in PyPSA, which is able to emulate various future scenarios for the UK grid and the subsequent changes to the optimal generation mixes and storage requirements. Hydrogen storage is modelled as LDES alongside lithium ion LFP battery storage, which covers short-term deficits. To capture the impact of path-dependent investment decisions, a myopic pathway optimisation was employed. This approach contrasts greenfield and brownfield approaches to evaluate how pre-existing generators influence deployment of hydrogen storage by 2050.

2 Methodology

2.1 Data Sourcing

In this study, a series of capacity and generation distribution optimisation simulations were conducted using PyPSA. These were modelled on 60 years of UK capacity factor data from 1960-2020 [7] with a 3 hour resolution. The model used a myopic pathway optimisation on an energy system with decennial planning years from 2020 to 2050. Technology costing, generation capacities and UK carbon dioxide emissions limits [8] were updated corresponding to the projected values in each planning year. Exponential decay was assumed when considering the 2040 limit as the fourth carbon budget is currently unset. The demand time series was not scaled to reflect future demand due to high level of associated uncertainty. All current and future technology costs were taken from the PyPSA technology-data resource [4]. These capital costs were annualised according to the technology lifetime.

The generation capacities were scaled for each generation technology type according to the NESO Future Energy Scenarios (FES) holistic transition pathway. The holistic transition pathway achieves net zero "through a mix of electrification and hydrogen" [9] and has the lowest levels of hydrogen dispatchable power. This was chosen as it is NESO's "central case" and a reasonable estimation for a lower bound utilisation of long duration hydrogen storage.

The energy storage modelled in the simulations included hydrogen storage combined with electrolyzers and hydrogen turbines, acting as long duration energy storage and lithium-ion-LFP battery storage, which serves as a short duration energy storage. The generators included oil, OCGT, CCGT, solar PV, biomass, nuclear, onshore wind and offshore wind, with oil and gas generators being the sources of carbon dioxide emissions. For every MWh of electricity generated, oil and gas generators were taken to produce 0.57 and 0.46 tCO₂, respectively. These were ob-

tained from stoichiometric relations where 44 GJ/tCH₂ 50 GJ/tCH₄ were converted to CO₂ emissions per MWh of electricity using generator efficiencies [7].

2.2 Single-Year and Multi-Year Optimisation

The model was initially built around a single-year time-span, with a minimum cost objective, which impacted the generation mix and storage required in each case. Models were developed for decennial single operating years from 2020 to 2050 using a greenfield approach. This was achieved by coupling historical demand and capacity factor data with carbon emissions constraints to define the relevant period. The generator and storage technologies had associated capital and marginal costs, informing their construction viability in the chosen period. The same approach was then applied to a 60-year model, aimed to evaluate hydrogen storage usage when accounting for multi-year time-spans. This reduced uncertainties compared to the single-year models, as inter-annual variability could be explored, capturing how consecutive years impacted the cost-optimal generation mix and subsequent potential of hydrogen deployment. For 60-year periods, all costs and carbon limits of the final year were extended to the entire time-span.

In addition to the carbon dioxide emissions limits, further constraints were added for each type of generation corresponding to the maximum installed capacity of each technology in the relevant operating year [6]. This prevented unrealistic biases being placed on individual generation sources and ensured a comprehensive model of the UK electricity network. These models were repeated using tank hydrogen storage instead of underground hydrogen storage to compare the required storage energy capacity in each case.

2.3 Myopic Pathway Optimisation

The simple myopic pathway optimisation model was developed such that each investment period was optimised without foresight of future costs. The model tracked the optimised capacity of previous periods in the pathway and used it as a base which was built on with the additional optimised capacity for the subsequent periods. The models in this study used decennial planning years from 2020 to 2050, with a model built for single operating years as well as a multi-year model, which used 10 operating years between each planning year. This allowed for a more cohesive and holistic progression towards 2050 in comparison to the non-pathway models as costing and emissions constraints are updated at each planning year. These pathway models were built to account for technology lifespans, so that decommission dates of each technology were tracked and the corresponding capacity was removed from the model once that date had been reached.

Results were obtained for both a brownfield and greenfield approach, offering alternative implications for the future development of the UK electricity network. The brownfield optimisation accounted for pre-existing installed generation capacities in 2010, which was then expanded upon to achieve cost-optimality while maintaining an acceptable environmental impact. The greenfield optimisation created a new system from 2020, which ensured that each period

had minimised costs while also adhering to the carbon dioxide emissions constraint. These were scaled relative to 1990 GHG emissions levels where 204 MtCO₂e [10] were emitted, 80% of which were assumed to be CO₂.

The myopic pathway optimisation used for the model allowed for a capacity expansion which was built on previous periods in the pathway series and allowed for the introduction of a decommissioning function. This was done for both a greenfield and brownfield approach. The brownfield approach used a build year for pre-existing capacity of 2010. This was used to provide a realistic representation of a recent UK electricity network, while giving the optimisation scope to create new generators before 2050. 2010 was chosen as the build year for the brownfield approach, as carbon budgets were first introduced in 2008-2012 for the UK [8]. Furthermore, coal was added to the 2010 pre-existing capacity for the brownfield approach. This was implemented to ensure demand was met, as coal made up a significant proportion of the 2010 generation mix. This coal capacity was subsequently decommissioned in 2020, as the final coal generation plant was decommissioned in 2024. These models were also repeated using tank hydrogen storage instead of underground hydrogen storage to compare the required storage energy capacities in each case.

3 Results and Discussion

3.1 Optimised Single-Year Model

This single-year model was optimised for minimising total cost and performed for decennial single operating years from 2020 to 2050, with a greenfield approach being used for each. These models conducted optimisation in isolation

for the costs and constraints for the corresponding year that was modelled, and were not part of a pathway optimisation.

The results using the 2020 annual carbon dioxide emissions limit of roughly 100 million tonnes per annum yielded a generation mix dominated by CCGT generation, which made up 54% of the total mix. This was expected due to the relaxed carbon dioxide emissions limit for 2020 and the low cost and high installed capacity of CCGT generation.

Oil and OCGT generation have very low contributions to the generation mix, with proportions of 0.40% and 0.36%, respectively. Due to the greenfield approach, oil and OCGT generation have no pre-existing installed capacity and are out-competed by CCGT generation. This was because the model favoured the significantly higher maximum installed capacity of CCGT compared to OCGT and the lower fuel cost of CCGT compared with oil generation. The extremely high investment cost of nuclear generation acted as a barrier to entry without pre-existing capacity.

Hydrogen storage was not used at all due to the high costs associated with it in 2020, and relatively non-constricting emissions limit. Battery storage was also not used by the model for this case, due to the high capacity of reliable, non-renewable generation, such that there were no short duration deficits which required mitigation.

The model was subsequently updated to reflect a single-year optimisation of 2030, 2040 and 2050, with projected costing, capacities and the emissions limits. The 2020 time series of capacity factors was used in each of these future, single-year models. This allowed for a prediction of the optimal unit capacities before increasing the complexity of the model by introducing multi-year time-spans.

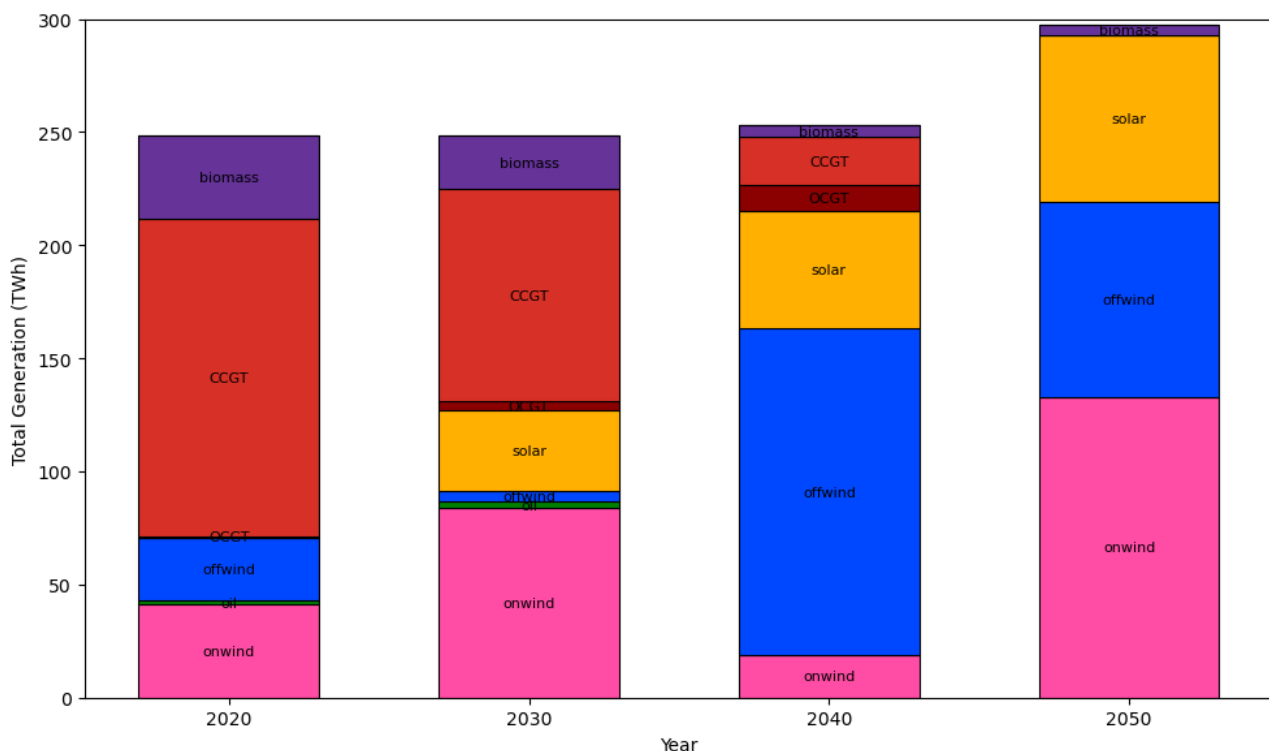


Figure 3.1a: Optimised Generation Mixes for Single-Year Models. This displays electricity mixes for isolated 1 year periods ending decennially from 2020 to 2050. CCGT usage decreases to zero and onshore wind dominates by 2050 due to its reduced capital cost.

As expected, the generation mix to meet the net zero emissions limit, shown in Fig. 3.1a, was biased further towards non-emitting generation technologies. As the emissions limit became stricter, leading up to 2050, the mix was comprised completely of non-emitting technologies. The projected costs associated with renewable and storage technologies decrease each decade, so hydrogen storage and battery storage became viable in the 2040 and 2050 models. In the 2040 model, the energy storage requirements for hydrogen and battery storage were 2.41 TWh and 52.1 GWh, respectively. In comparison, the 2050 model had significantly greater hydrogen and battery storage requirements

of 9.93 TWh and 78.5 GWh, respectively, due to the decreased costs. The 2040 generation mix used a larger proportion of offshore wind, 57.0%, which is more expensive, but more reliable than onshore wind generation which made up 7.47% of the mix. This was necessary due to the lower hydrogen storage capacity in this model. In the 2050 model, as hydrogen storage costs were reduced, the model found it more optimal to use an increased hydrogen storage capacity in tandem with a greater proportion of the cheaper, more seasonal onshore wind generation, which made up 44.7% of the mix, with offshore wind generation at 29.0%.

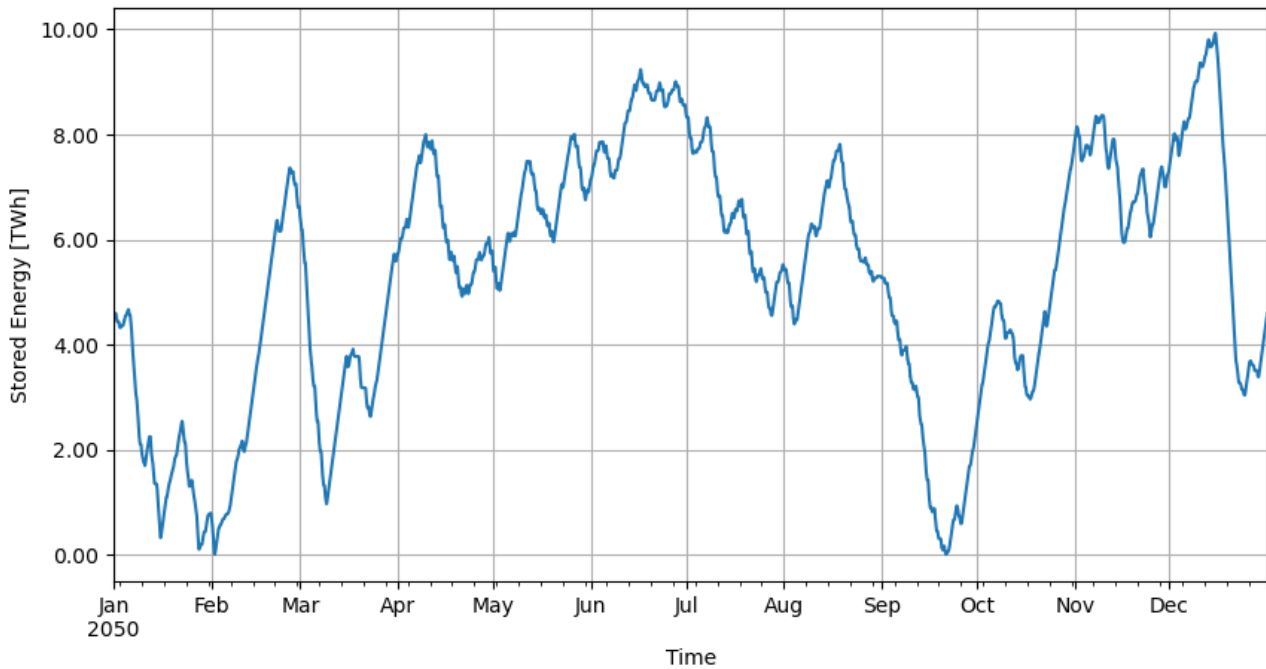


Figure 3.1b: 2050 Single-Year Underground Hydrogen Storage Filling Level. This displays a seasonal pattern, with a summer peak and a winter peak. The winter peak (9.93 TWh) is greater, due to the high dependence on onshore wind generation. This is because onshore wind generation is highly seasonal and is at its greatest in the winter months.

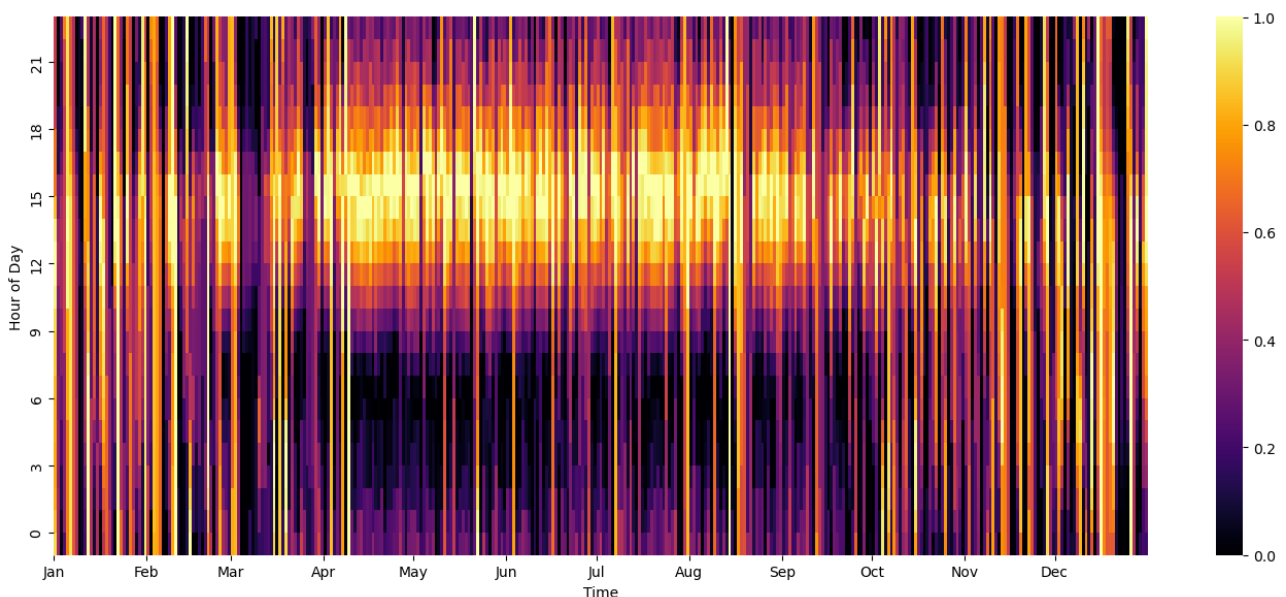


Figure 3.1c: 2050 Normalised Single-Year Storage Filling Level Heat Map: Li-ion-LFP Battery Storage. This displays a diurnal pattern with increased charging and discharging across the summer months due to the lack of curtailment modelling.

The time series for hydrogen storage filling level in both the 2040 and 2050 model showed a clear seasonal pattern, shown for the 2050 model in Figure 3.1b, in tandem with a diurnal pattern for battery storage, shown for the 2050 model in Figure 3.1c. This further emphasised their difference in suitability, depending on the deficit duration. Figure 3.1c shows the battery storage was charged and discharged frequently over the summer months. This was due to lack of curtailment in the model. The high surplus of generation in the summer months resulted in the battery storage effectively acting as curtailment to prevent flooding of the hydrogen storage.

Figure 3.1b shows there was an increase in the hydrogen storage filling level over the summer months, as the demand was lower and solar generation was at its peak. Another significant peak was in December, which corresponded to the period of highest wind generation, which the models relied heavily on. In the 2050 model, this winter peak was greater than the summer peak. The winter peak was slightly less pronounced for the 2040 model, being equal in magnitude to the summer peak, as it used a greater proportion of the less seasonal offshore wind generation compared to 2050. Throughout the models in this study, battery storage displayed a greater power capacity requirement compared to hydrogen storage. For this 2050 model the power capacities were 40.0 GW for battery and 18.1 GW for hydrogen storage. The greater power requirements of batteries can be explained by the faster charge-discharge cycles exhibited, with cycling periods lasting just 5.6 hours [11]. In addition, this reflects the slow charging and discharging of hydrogen storage due to the conversion steps.

When repeated with tank hydrogen storage, the 2040 and

2050 models showed a significant decrease in the maximum hydrogen storage capacity for 2040 and 2050, which were 111 GWh and 5.58 TWh, respectively. The generation mixes in 2040 and 2050 differed depending on whether the model used underground or tank hydrogen storage. As underground hydrogen storage has lower costs per unit of stored energy, the model was able to use a higher proportion of onshore wind, which is more seasonal but cheaper than offshore wind. For the 2050 model this was 44.7% onshore wind generation with underground storage compared to only 3.21% with tank storage. Conversely, when the model used tank hydrogen storage, as storage costs were more expensive, the model favoured using a much greater proportion of offshore wind, increasing to 70.5% from 29.0% in the 2050 model, which is a more consistent source of generation.

3.2 Optimised 60-Year Model

At this stage of the study, following previous literature, the time series was expanded to use the entire capacity factor dataset and demand data from 1960-2020. This time series was used to model decennial 60 year time-spans, with final years of 2020, 2030, 2040 and 2050, by using the costing, capacity and annual emissions limit corresponding to the final year in the series. Another simplification was the application of the 1960-2020 demand and capacity factor time series to each period. This was used as demand and weather-dependent generation availability is uncertain for future years. Using the full dataset for a multi-year time-span allows the model to account for extreme weather events and the inter-annual relationships necessary to provide a robust optimisation of storage utilisation in the UK grid.

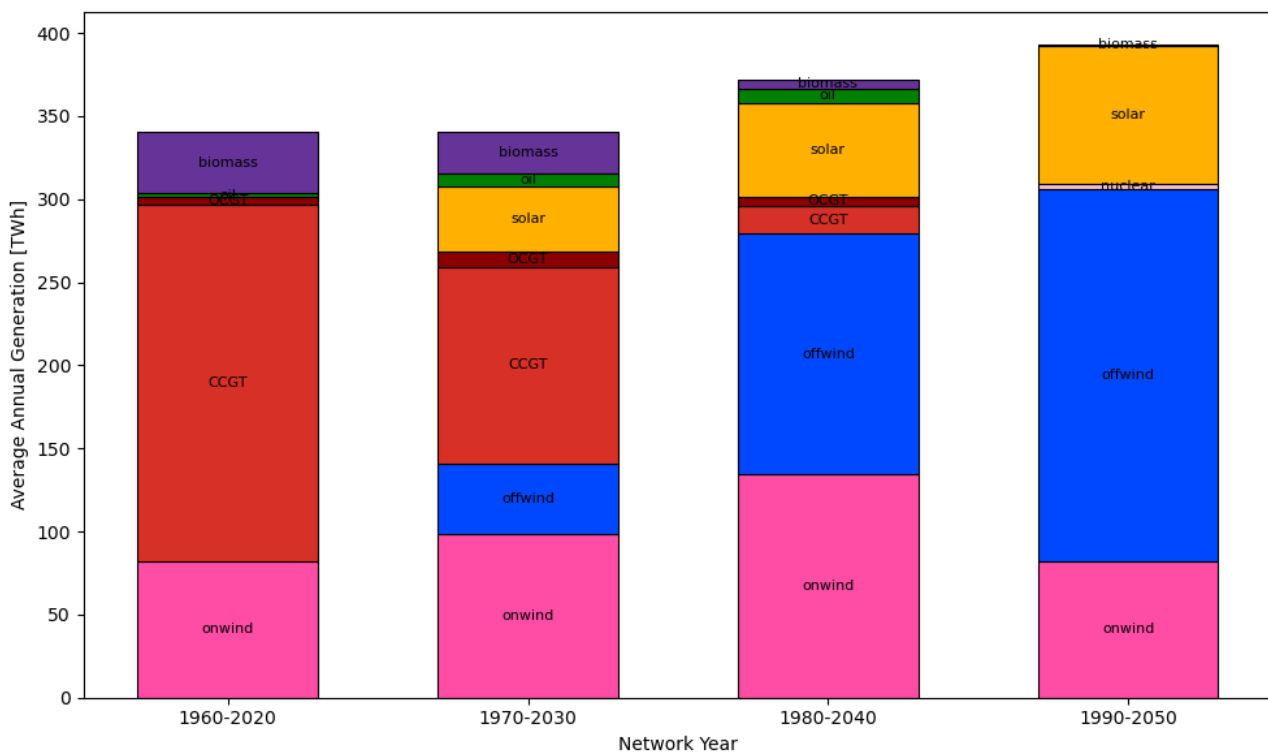


Figure 3.2a: Optimised Generation Mixes for Multi-Year Models. This displays electricity mixes for isolated 60 year periods ending decennially from 2020 to 2050. CCGT usage decreases to zero and offshore wind dominates by 2050 due to its increased reliability.

Once again, it was found that hydrogen storage was only used for the 2040 and 2050 models. For the models in Figure 3.2a, the generation mix was dominated by offshore wind generation, which was expected for the UK due to its geographical location in north-west Europe. This greater proportion of offshore wind compared to onshore wind generation was the major difference between the single-year

and multi-year model generation mixes. This was due to the increased reliability of offshore wind generation, which is crucial for robustness in meeting demand for multi-year studies. Therefore, for multi-year cases, it is seen as worthwhile to incur the larger initial capital investment for the increased and more consistent generation obtained through offshore wind farms.

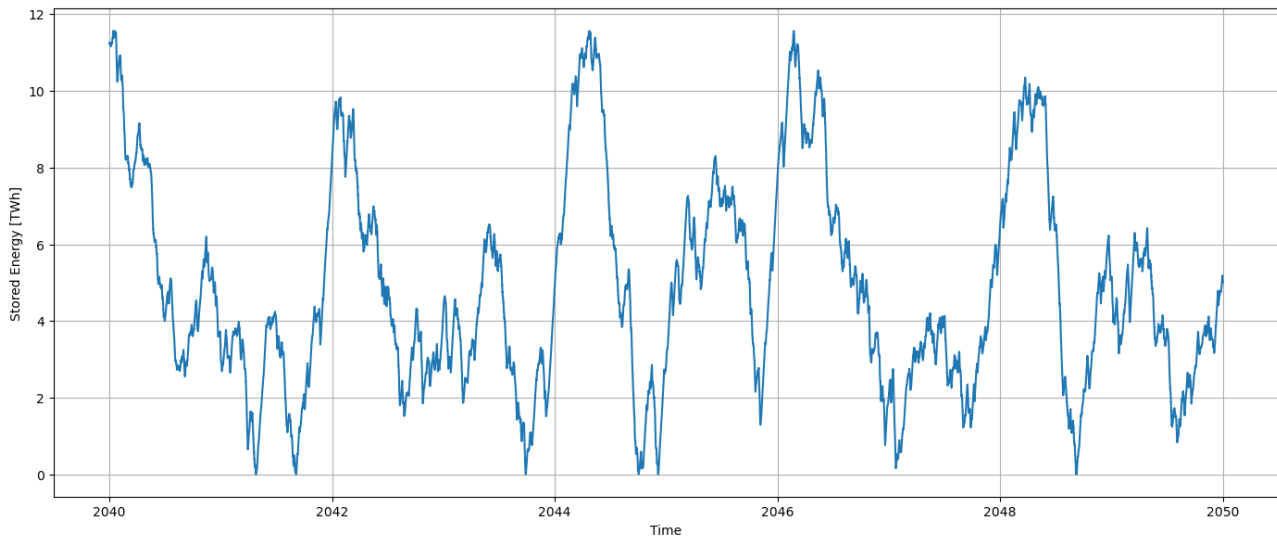


Figure 3.2b: 2040-2050 Multi-Year Underground Hydrogen Storage Filling Level. The profile shows significant inter-annual variability in dispatch. The required storage is 11.5 TWh, greater than the capacity for a single year. This demonstrates the requirement for robustness of LDES over extended periods, where it completely discharges 6 times over a decade to maintain electricity supply.

The hydrogen storage filling level in Figure 3.2b shows frequent discharging when necessary to meet demand. The energy level peaks occur with less seasonality than in the single year model, emphasising the generation and demand differences between years. The multi-year captures the significant variation in energy surplus and demand, ensuring a robust build with the years of reduced generation able to be covered. The hydrogen energy capacity for this multi-year model was 3.86 TWh and 11.5 TWh for the model ending in 2040 and 2050, respectively. This was an increase on the single-year model requirements, which had corresponding energy capacities of 2.41 TWh and 9.93 TWh for 2040 and 2050, respectively. The greater hydrogen energy storage capacity required by the multi-year model emphasised the need for increased network robustness when considering multi-year time-spans.

3.3 Myopic Pathway Optimisation for a Single-Year Temporal Domain

The brownfield and greenfield generation mixes shown in Figure 3.3a exhibited trends between 2020 and 2040 which displayed a clear decreasing use of CCGT, a dominating source in the 2020 planning year. This was expected as the model adhered to the changing emissions constraints applied to each period in the optimisation. The greenfield approach showed no nuclear installation at all, due to the high investment costs associated with the construction of a new power plant. This contrasts the brownfield case in which the

model incurred no additional capital cost when using 2010 generators. Therefore, nuclear was used until 2050, the end of its 40 year lifetime, though with decreasing total generation as the model did not find constant purchase of nuclear fuel optimal. Furthermore, there was minimal total surplus generation compared to demand when summed over each of these years. In 2050, total generation exceeded annual demand by 5.71%, whereas the surplus remained below 1.05% throughout 2020–2040. This can be explained by the hydrogen storage utilisation, with an RTE of 40%. The consequential losses from electrolyzers and turbines mean that over-production from generators in times of surplus is required.

The 2050 planning year was dominated by weather-based renewable generation, primarily from offshore and onshore wind, which was expected for the UK due to its geographical location being an island in the north-west of Europe. The UK climate permits increased wind generation and efficiencies, making wind-based generators inherently favourable. Unlike 2020-2040, this planning year utilised hydrogen storage and required over-generation in periods of energy surplus, explaining the notably increased total generation. Despite solar generators accounting for a small portion of the generation mix, their installed capacity was significantly greater than offshore or onshore wind in Table 3.4a. This is because solar panels have much lower capacity factors for the UK, meaning larger scale deployment is required to yield a similar output.

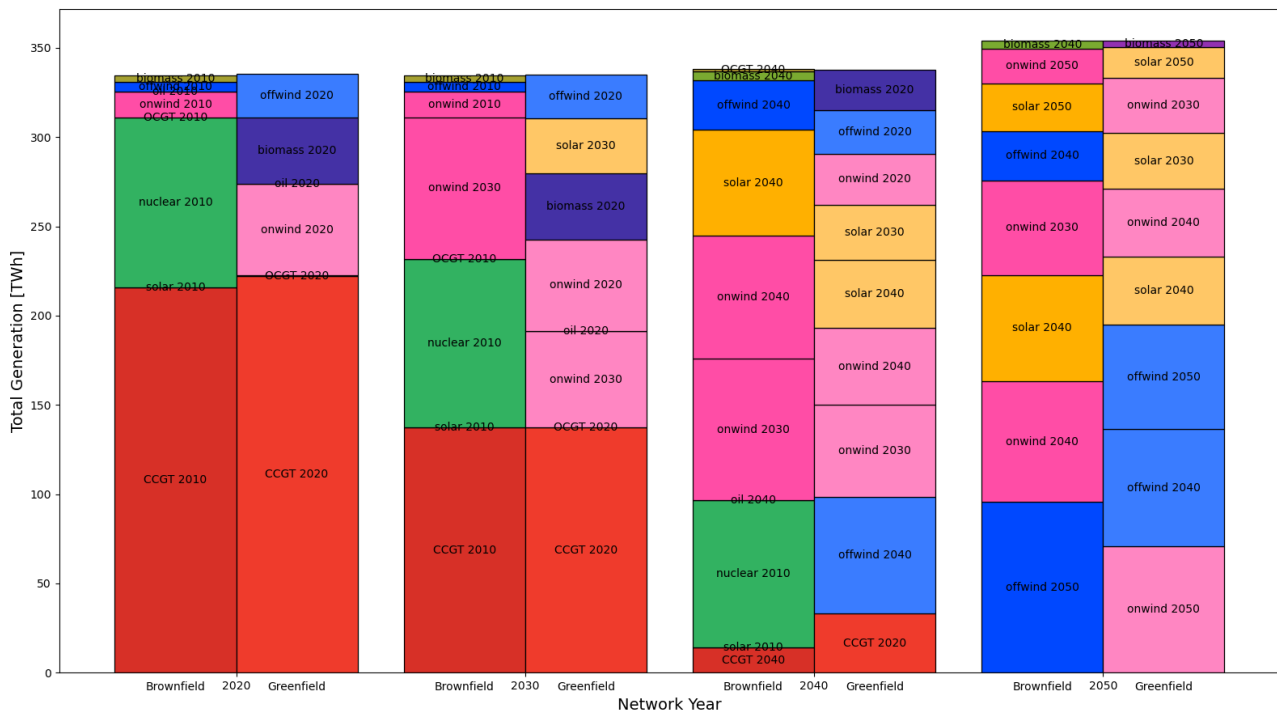


Figure 3.3a: Brownfield and Greenfield Single-Year Myopic Optimisation Generation Mixes (Underground Storage). Again CCGT usage decreased between each year and nuclear is consistently used in the brownfield model until 2040. The 2050 mix is dominated by wind generation, leveraging the UK's geographical advantage for high yields of onshore and offshore wind over solar power.

Table 3.3a: Single-Year Myopic Optimisation Generator Installed Capacities by Technology (MW). All increases in total capacity are from newly built generators and all decreases are from decommissioning. The values are bounded by the NESO FES holistic transition.

Technology	2020		2030		2040		2050	
	BF	GF	BF	GF	BF	GF	BF	GF
CCGT	30,900	30,900	20,600	20,600	5,410	5,410	0	0
OCGT	1,170	1,800	1,170	1,800	2,460	0	0	0
Biomass	431	4,260	431	4,260	951	4,260	951	833
Nuclear	10,900	0	10,900	0	10,900	0	0	0
Offwind	1,340	5,850	1,340	5,850	6,660	21,700	29,800	30,000
Oil	5,710	212	5,710	212	140	212	140	0
Onwind	4,080	14,600	26,500	29,800	42,000	42,000	47,500	47,500
Solar	95.0	0	95.0	34,800	66,900	77,700	97,000	97,000

Note: BF = Brownfield, GF = Greenfield

In the single-year myopic pathway optimisation, underground hydrogen storage was not deployed until 2050. This was a key difference from the earlier, isolated models which did not use pathway optimisation and had hydrogen storage installed in 2040 and 2050. This difference was due to the myopic nature of these pathway models, meaning it had no foresight of future costs. This meant that generation capacity was built up significantly enough in the previous planning years due to the high storage costs, that hydrogen storage was not required in 2040, despite the reduced storage costs.

The storage requirements for hydrogen storage in 2050 in the brownfield approach pathway model were energy and power capacities of 2.51 TWh and 12.6 GW, respectively. The greenfield results were very similar with minor differences, shown in Figure 3.3b. The myopic pathway optimisation was also repeated using tank hydrogen storage, to

compare the effectiveness of both storage types under identical conditions. This resulted in energy and power capacities of 1.74 TWh and 10.2 GW, lower than the requirements for underground storage, as utilisation is less favourable due to increased storage costs.

The filling level profile in Figure 3.3b displayed significant seasonality, similar to the single-year optimisation in Figure 3.1b. The pattern of the hydrogen storage retaining charge over the summer months before dispatch around the third quarter closely follows previous models. Again, in a similar fashion, the storage replenished capacity towards the extremes of each calendar year due to increased wind generation over the winter months. The maximum single-year myopic energy capacity requirement was less than the single-year model due to the presence of capacity built in previous years, reducing the necessity of hydrogen storage.

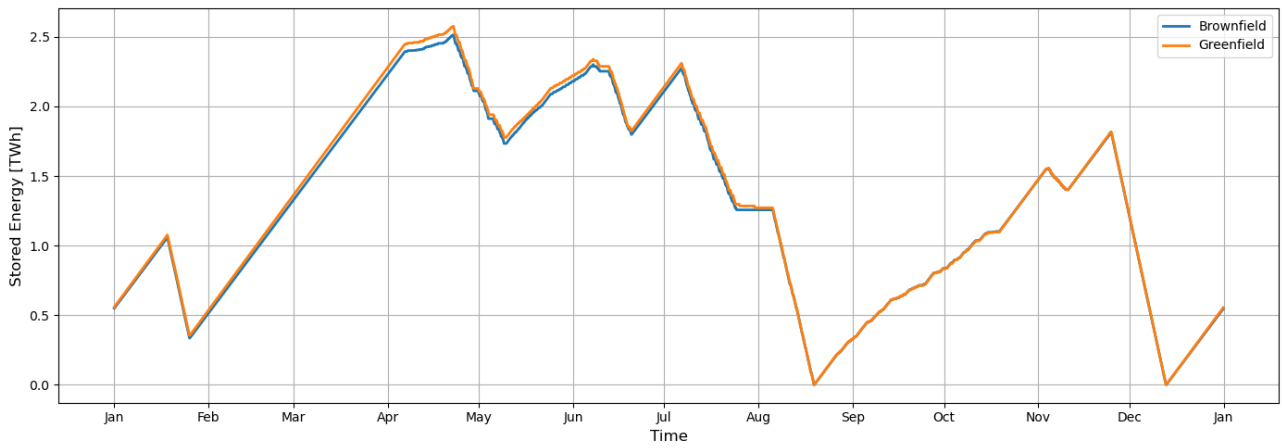


Figure 3.3b: Single-Year Myopic Optimisation Hydrogen Storage Filling Level. This shows strong seasonality, with brownfield and greenfield displaying similar patterns and magnitudes. There is a maximum brownfield energy storage requirement of 2.51 TWh.

3.4 Myopic Pathway Optimisation for a Multi-Year Temporal Domain

The final stage of this study was to perform the myopic pathway optimisation for a multi-year temporal domain. This combined the complexity of inter-annual relationships with both brownfield and greenfield approaches to the myopic pathway optimisation. The pathway optimisation allowed for the analysis of a more realistic progression towards 2050 and the net-zero emissions goal. Observations of the effects this has on the optimal utilisation of hydrogen storage to mitigate seasonal generation deficits were then made.

The multi-year myopic optimisation displayed the same general trends in the generation mix as the single-year myopic optimisation for both the brownfield and greenfield approaches, as shown in Figure 3.4a. The most significant dif-

ferences were in the 2050 planning year, where there was a much larger proportion of offshore wind generation and no additional onshore wind generation capacity installed in 2050 for the multi-year model. This was due to the significantly higher and less seasonal capacity factors of offshore wind which were required for robustness in a multi-year model. Although the offshore wind projected investment costs for 2050 remained higher than those for onshore wind, the costing gap had dropped enough so that the model found an increased offshore wind proportion more optimal due to its greater capacity and reliability. These results reflect the UK's plans to invest more heavily in offshore wind generation. According to the NESO holistic transition pathway [6], offshore and onshore wind installed power generation capacities in 2050 are set to be 104 GW and 47.5 GW respectively.

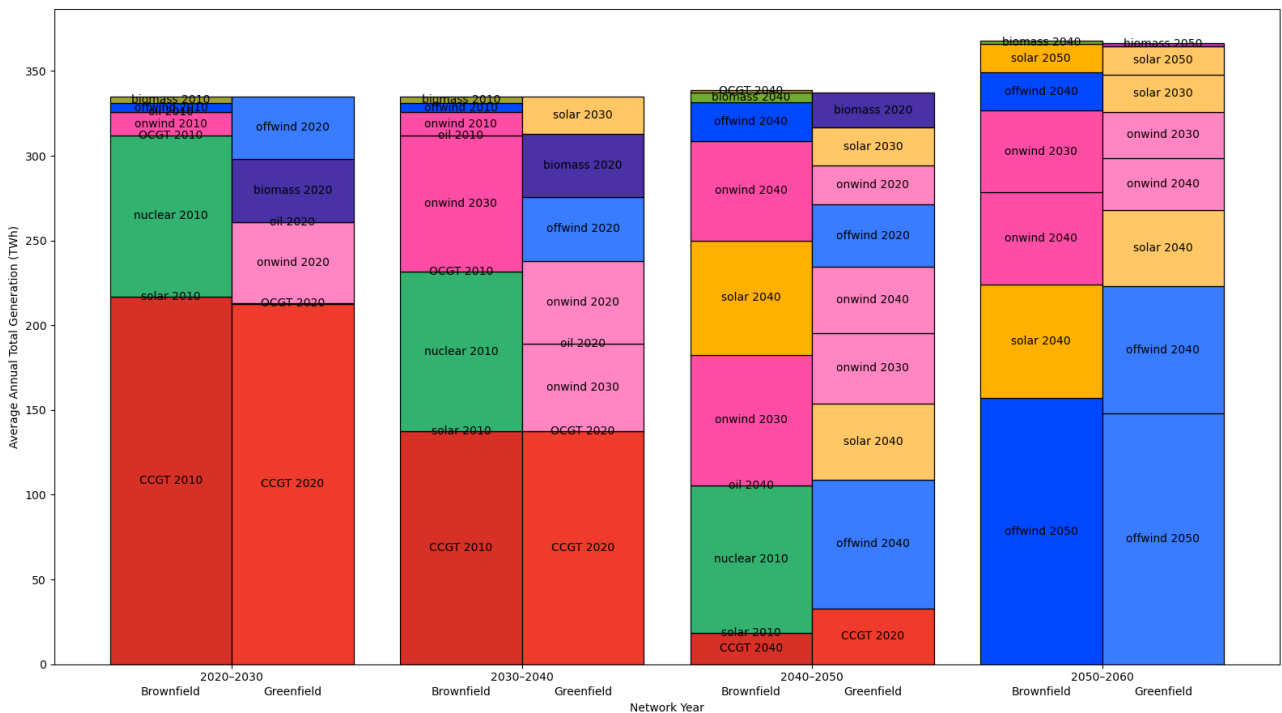


Figure 3.4a: Brownfield and Greenfield Single-Year Myopic Optimisation Generation Mixes (Underground Storage). CCGT usage decreased between each year and nuclear is consistently used in the brownfield model until 2040. The 2050 mix has a greater proportion of offshore wind than the single-year model, due to the importance of greater reliability and robustness for a multi-year model.

Table 3.4a: Multi-Year Myopic Optimisation Generator Installed Capacities by Technology (MW). All increases in total capacity are from newly built generators and all decreases are from decommissioning. The values are bounded by the NESO FES holistic transition.

Technology	2020		2030		2040		2050	
	BF	GF	BF	GF	BF	GF	BF	GF
CCGT	30,900	30,900	20,600	20,600	5,410	5,410	0	0
OCGT	1,170	1,800	1,170	1,800	2,460	0	0	0
Biomass	431	4,260	431	4,260	951	4,260	951	833
Nuclear	10,900	0	10,900	0	10,900	0	0	0
Offwind	1,340	9,360	1,340	9,360	5,800	28,700	46,200	57,400
Oil	212	212	140	140	140	0	0	0
Onwind	4,080	14,600	27,900	29,800	42,000	42,000	42,000	27,400
Solar	95.0	0	95.0	25,600	77,700	77,700	97,000	97,000

Note: BF = Brownfield, GF = Greenfield

For the same reasons described in Section 3.3 for the single-year myopic pathway optimisation model, hydrogen storage was not viable in the multi-year myopic pathway optimisation model until the 2050 planning year, regardless of the starting network. The brownfield multi-year myopic pathway optimisation model was the most accurate model developed for this study. However, the results from the previous, isolated models discussed in Section 3.1 and 3.2 had hydrogen storage installed in 2040. It is likely that hydrogen storage would have been deployed within the 2040-2050 decade if the planning year resolution was improved. Figure 3.4b shows the hydrogen storage filling level across the 2050-2060 decade. The pattern exhibited here agreed closely with the isolated multi-year model for 2050, discussed in Section 3.2. This multi-year model had a 4.40 times greater maximum storage requirement and a 3.24 times greater storage energy level for the 2050 year summer peak compared to

the single-year myopic pathway model. This result was significant and emphasised the importance of multi-year modelling to develop a robust network, as in the absence of inter-annual relationships, single-year models can produce an under-estimation in the LDES requirements. Furthermore, this pathway optimisation was repeated using tank hydrogen storage rather than underground hydrogen storage. The maximum storage requirements in this case were 5.35 TWh and 5.15 TWh for the brownfield and greenfield approach, respectively. Again, as with the single-year myopic model, underground storage was utilised to a greater extent due to its lower cost per unit of energy stored. The brownfield multi-year myopic optimisation model which used tank storage had a total network cost in 2050 which was 17.0% greater than the equivalent model which used underground hydrogen storage.

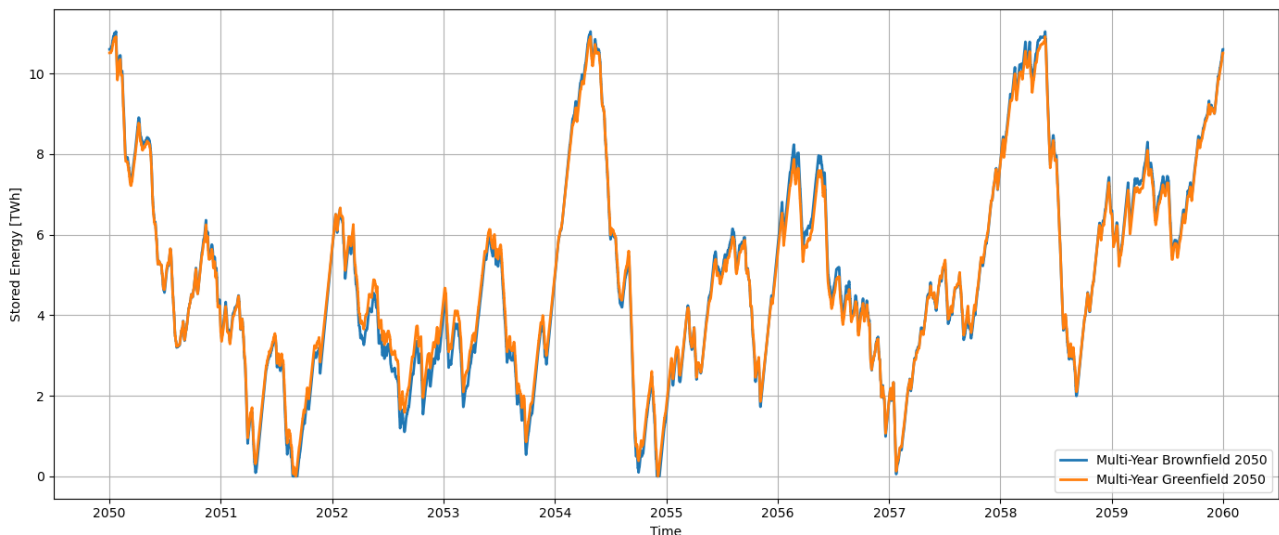


Figure 3.4b: Multi-Year Myopic Optimisation Hydrogen Storage Filling Level. This shows significant inter-annual variability, with brownfield and greenfield displaying similar patterns and magnitudes. There is a maximum energy storage requirement of 11.0 TWh.

4 Conclusions

Hydrogen's viability for LDES in the UK has been investigated through a series of increasingly complex optimisation models, designed for a comprehensive overview of different time-spans. This included greenfield optimisation for single and 60 year periods, followed by a brownfield and greenfield approach to myopic pathway optimisation, aimed to

depict a more realistic energy transition pathway from 2020 to 2050. The brownfield network replicated the 2010 UK electricity network, as this fell within the time period of the first carbon budget and provided an accurate base to examine the transition to 2050. It has been proposed that hydrogen storage will become more competitive than thermal storage after 2045 for LDES [12]. This agrees with results in this paper as all models displayed consistency in their

predictions for the hydrogen storage introduction timeline, indicating incorporation of hydrogen LDES by 2050.

In the single-year model, hydrogen appeared by 2040 because the system was optimised for that year alone, without considering earlier or later years. In the multi-year model, hydrogen also appears around 2040; however, this model applied technology costs and constraints from the final year (2040) uniformly across all years. As a result, the model may introduce hydrogen earlier than a realistic gradual cost evolution would suggest. Additionally, the myopic models exhibited gradual increases in installed capacities without foresight of costs, meaning that by 2040, the installed generators were adequate to meet demand without reliance on hydrogen LDES. If rather than decennial planning years, the resolution was improved to five-year periods, hydrogen may also be utilised before 2050.

The models developed throughout this report further emphasised the importance of multi-year modelling when analysing the role of LDES in the UK grid. The results consistently showed a significant increase in the hydrogen storage energy capacity for the multi-year models over those with single operating years. For the most complex model in this study, the brownfield myopic pathway optimisation, the multi-year model had a 4.4 times greater hydrogen energy storage capacity over the corresponding single-year model.

Furthermore, underground hydrogen storage was compared against tank storage, with the former exhibiting greater deployment under identical conditions due to its lower costs per unit of stored energy. This attribute results in a greater scalability, such that underground hydrogen storage is more suitable for the UK's seasonal and national-scale storage requirements.

5 Outlook

The myopic pathway optimisation models would be enhanced by using an improved resolution for the planning years. This would provide a more accurate representation, as costs and installed capacities would be updated more frequently. The brownfield approach could also be updated to better reflect reality. The current implementation of the brownfield approach was heavily simplified, as it used a single build year, so that within the model all pre-existing capacity was installed in 2010.

Future studies can expand upon the current models, increasing their complexity and realism by accounting for factors which would affect the viability and extent of hydrogen deployment. This would include the addition of other forms of storage, such as pumped hydro storage (PHS), thermal energy storage (TES), and adiabatic compressed air energy storage (ACAES). Although these forms of energy storage are less ideal for truly seasonal applications, they would provide multi-day to multi-week storage to the model. These may be implemented before hydrogen storage and would affect the charge-discharge cycle of hydrogen storage.

Accounting for generation and storage locations with transmission costs and losses would directly impact underground hydrogen storage, due to its inflexibility in location. This would potentially allow more modular and relocatable stor-

age options such as TES and tank hydrogen storage to become more viable. Furthermore, inclusion of curtailment and international transmission would improve the realism of the model and impact the storage requirements. In certain cases, during a surplus, it may be more economically optimal to curtail or export energy internationally, and during a deficit, it may be possible to import energy, reducing the storage requirements. Curtailment also prevents storage flooding, which would allow renewable generation capacity to increase without having to increase storage capacity.

Acknowledgments

The authors would like to thank Ebbe Kyhl Gøtske and Martin Stringer for their continued support and guidance throughout the completion of this research.

References

- [1] Department for Energy Security and Net Zero. Digest of UK Energy Statistics (DUKES) 2025. Technical report, HM Government, 2025.
- [2] John-Joseph Marie. Faraday Insight 21: Batteries in Stationary Energy Storage Applications. Technical report, The Faraday Institution, 2024.
- [3] House of Lords Science and Technology Committee. Long-duration energy storage: get on with it. Technical report, House of Lords, 2024.
- [4] Fabian Neumann, Elisabeth Zeyen, Tom Brown, et al. PyPSA/technology-data: v0.9.2, 2024.
- [5] Ruhnau, Oliver and Qvist, Staffan. Storage requirements in a 100% renewable electricity system: Extreme events and inter-annual variability. *Environmental Research Letters*, 2022.
- [6] National Energy System Operator. Future Energy Scenarios (FES) 2024. Technical report, NESO, 2024.
- [7] Ebbe Kyhl Gøtske, Gorm Bruun Andresen, Fabian Neumann, and Marta Victoria. Designing a sector-coupled European energy system robust to 60 years of historical weather data. *Nature Communications*, 2024.
- [8] Department for Energy Security and Net Zero. Guidance: Carbon budgets, 2024.
- [9] National Energy System Operator. Future energy scenarios: Pathway to net zero. Technical report, NESO, 2025.
- [10] Department for Energy Security and Net Zero. Powering Up Britain: Net Zero Growth Plan and Carbon Budget Delivery Plan – Analysis Methodology Technical Annex, 2023.
- [11] Marta Victoria, Kun Zhu, Tom Brown, Gorm Bruun Andresen, and Martin Greiner. The role of storage technologies throughout the decarbonisation of the sector-coupled European energy system. *Energy Conversion and Management*, 2019.
- [12] Li Zheng, Du Binglin, Nils Petersen, Liu Pei, and Manfred Wirsum. Potential of hydrogen and thermal storage in the long-term transition of the power sector: A case study of China. *Energy*, 2024.

Optimizing Copper (Cu) Electrodeposition via Different Control Strategies

Xin Chen and Yi Li

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

With the steady increase in copper demand and heightened awareness of the environmental impact of heavy-metal discharge, electrodeposition has emerged as a sustainable and efficient approach for recovering copper from diverse waste solutions. The goal of this study is to optimize the Cu electrodeposition process from acidic aqueous waste in two main pathways by adjusting the dynamic optimal electrode potential and changing the flow rate, to meet the regulatory discharge concentration target of 1 mg/L. [1] Several short-period batch experiments were conducted to construct a calibration curve for optimum potential as a function of electrolyte concentration. Optimal potential was identified corresponding to highest Faradaic efficiency (FE) and lowest specific energy cost (SEC). This research further evaluated the relationship between current density and applied potential, discovering a distinct four-region phenomenon. Kinetic and mass-transfer transitions and limitations during deposition were clarified. Long-period batch experiments presented two control strategies both compared with the conventional constant potential operating mode. By examining the reaction's intrinsic kinetics, applied potential, copper concentration and key performance metrics including FE, recovery and SEC, the results demonstrated that the dynamic potential approach achieved the lowest SEC (average cost of £ 583 per tonne of Cu), while operation at an elevated flow rate yielded the highest copper recovery (98.9 %) thanks to the mass transfer enhancement, albeit with diminished FE in later period. Corresponding to these findings, integrating dynamic potential control with increased flow rate is recommended as a more efficient and cost-effective electrodeposition operating mode.

1. Introduction

With the rapid development of electrochemical industries such as electroplating, metallurgy and printed circuit board manufacturing, [2] large amount of wastewater rich in copper ions has been generated, posing environmental risks. As a harmful heavy metal, copper could significantly threaten aquatic ecosystems if directly discharged into natural water bodies. Adverse consequences may include impaired fish gills, hindered microorganism growth, and reduced biodiversity. [3] Human health could also be affected through drinking water and agricultural products that contain trace amounts of copper. Potentially, this could lead to problems such as cardiovascular disease, cognitive deterioration and weakened immune systems. [4] Economically, copper is frequently used in the green transition, particularly in renewable energy storage and transportation. [5] However, the dramatic rise in copper demand has transcended the supply of refined copper ore, highlighting the importance of recovering copper from industrial wastewater. Copper recovery thus offers dual environmental and economic benefits.

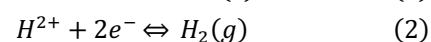
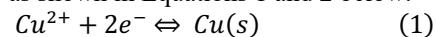
This study aims to optimize copper electrodeposition using a two-stage approach. First, the relationship between copper concentration and optimal applied electrode potential in an electrochemical reaction needs to be established through short-period experiments. Then, based on previous correlation, the feasibility of electrodeposition process via dynamic potential control methodology for a long-term experiment can be evaluated. This work hypothesizes that by dynamically adjusting applied potential according to real-time copper concentration, high efficiency electrodeposition regime could be prolonged compared to the conventional fixed-potential strategy, thereby shortening the treatment time and lowering energy

consumption. The effect of electrolyte flow rate on mass transfer is also examined.

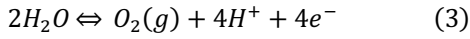
2. Background

Since the industrial development, the discharge of large amounts of non-biodegradable heavy metals such as copper into waterways has surged dramatically, posing significant environmental and human health impacts. [6] This drives the necessity for effective wastewater treatment to meet the regulated copper discharge limit, 1 mg/L, according to the Environmental Protection Agency (EPA). [1] There is a range of available methods to lower copper concentration, involving chemical precipitation, ion exchange, adsorption, and electrodeposition. [7] However, chemical precipitation relies on a large amount of reagents and is limited by unstable sludge production, which is only suitable for small plants. [7] Ion exchange and adsorption techniques require long pretreatment and regeneration of resins and absorbents, while the absorbent itself can be hazardous as well. [8] Among existing methods for wastewater purification, electrodeposition as a direct metal recovery method attracts great attention and appears to be one of the greenest and most efficient approaches.

The fundamental process of electrodeposition takes place in an electrolytic cell, which consists of a copper-based waste solution (electrolyte), a cathode, and an anode. Upon applying a voltage, redox reactions are driven electrochemically. At the cathode, Cu^{2+} ion is reduced to metallic copper, and hydrogen evolution side reaction occurs as shown in Equations 1 and 2 below.



Simultaneously, the anode reaction is typically water oxidation to oxygen gas, as shown in Equation 3 below.



Copper and its compounds participate in multifarious electrochemical reactions, and each reaction is characterized by a distinct standard potential with respect to the standard hydrogen electrode (SHE). These reactions are contingent on both pH and potential, as described by the Nernst equation, a Pourbaix diagram (potential-pH diagram, shown in Figure 1) could be constructed to integrate all electrochemical reactions related to copper together in a single diagram. [9] Such a diagram could be employed to interpret thermodynamic stability regions for different copper species in aqueous environments.

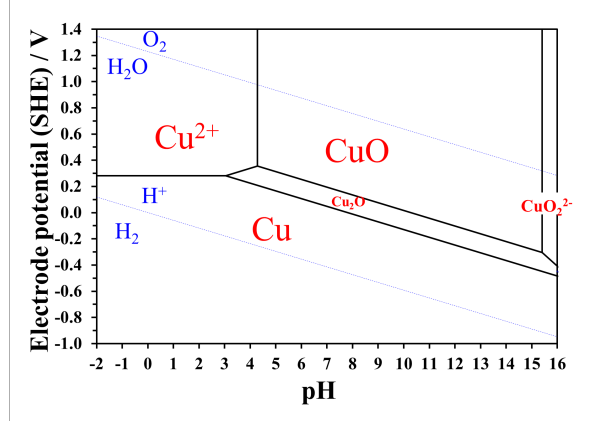


Figure 1 Pourbaix diagram for a Cu(II)-H₂O system at 298 K, $p(O_2) = p(H_2) = 1 \text{ atm}$ and total dissolved Cu(II) activities of 0.24 M

The reduction of copper (II) ion to pure copper metal is the main interest in this study. To avoid the unwanted formation of copper oxides, all electrodeposition experiments need to be carried out under strong acidic conditions (pH less than around 2.3). Moreover, the thermodynamic driving force for these reactions is governed by the Nernst equation shown below, which defines the equilibrium potential.

$$E_{Cu^{2+}/Cu}(SHE) = E_{Cu^{2+}/Cu}^0(SHE) + 0.0296 \cdot \log [Cu^{2+}] \quad (4)$$

Where $E_{Cu^{2+}/Cu}^0(SHE)$ is the standard potential, and $[Cu^{2+}]$ is the copper concentration. This equilibrium potential presents as a horizontal line in the acidic region of the Pourbaix diagram (Figure 1), which can be utilized to approximate the applied overpotential in the copper recovery process for a certain concentration. At fixed concentration, if a more negative potential than the equilibrium value is applied (i.e. a negative overpotential), the reaction will be driven into net reduction and hence favors the deposition of elemental copper. The Pourbaix diagram also includes the oxidation and reduction of water, which are intrinsic to aqueous systems, particularly at low copper concentrations. Water acts as a competing factor during copper electrodeposition. To elaborate this, if excessively high overpotential is applied, hydrogen evolution may become dominant over copper reduction, wasting energy and decreasing the overall efficiency.

Despite its advantages, the traditional electrodeposition recovery method is restricted by the limitation of mass transport, while also maintaining high-quality recovery

of copper, especially for dilute solutions; therefore, there is always a tradeoff between recovery rate, cost, and energy consumption. To evaluate and optimize the overall efficiency, Faradaic efficiency (FE), and specific energy consumption (SEC) are introduced as key metrics and quantified employing the equations shown below.

$$FE = \frac{m_{exp} \cdot n \cdot F}{Q \cdot M_{metal}} \quad (5)$$

where m_{exp} is the experimentally determined mass of the deposit, n is the number of transferred electrons, F is the Faraday constant, Q is the charge passed during electrodeposition and M_{metal} is the molecular weight of the deposited metal.

$$SEC = \frac{Q \cdot V}{FE \cdot \Delta m} \quad (6)$$

where V is the cell potential difference (i.e. between anode and cathode) and Δm is the theoretical mass change.

3. Method

3.1. Set up

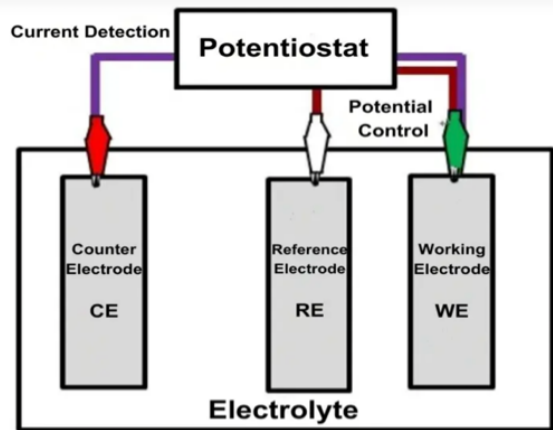


Figure 2 The schematic image of the three-electrode setup that enabled control of the Cu deposition potential on the cathode

In this study, all the copper electrodeposition experiments were carried out in a bench-scale reactor operating in batch-recycle mode. A three-electrode configuration was utilized as it allows isolated study of the cathode (copper reduction) performance. [10] The schematic image of three-electrode setup is shown in Figure 2 [11] below. Nickel foam was employed as the working electrode while Ti|IrO_x and AgCl|Ag were used as counter and reference electrode, respectively. A waste solution containing acidic copper sulfate salt and trace amounts of sodium sulfate was used as electrolyte. All experiments were conducted under chronoamperometry mode (i.e. fixed applied potential, while current is measured) via NOVA software, which is responsible for program settings and data collection.

3.2. UV-vis spectroscopy

3.2.1. Mechanism of UV-vis spectroscopy

Ultraviolet-visible (UV-vis) spectroscopy was employed to quantify copper concentrations. The technique is based on the absorption of specific wavelengths of light by copper ions, which occurs

when the photon energy exactly matches the energy difference between electronic energy levels. Absorbance is concentration dependent as described by the Beer-Lambert Law (Equation 7), where A is the light absorbance, ϵ is the molar absorptivity, ℓ is the thickness of cuvette (10mm cuvette was used in this study) and C is the concentration of the sample. [12]

$$A = \log_{10} \left(\frac{I_0}{I} \right) = \epsilon \cdot \ell \cdot C \quad (7)$$

Measurements were performed using a Shimadzu UV-2600 double-beam spectrophotometer, scanning from 200 to 1100 nm. A reference cell containing 1% nitric acid was used to correct the absorbance readings of the sample solutions.

3.2.2. Copper concentration calibration

Five calibration solutions (concentrations are 0.001 M, 0.005 M, 0.01 M, 0.05 M, 0.1 M) were prepared by dissolving different amounts of copper sulfate pentahydrate crystal into 1% nitric acid solution. The mass required for each solution was calculated using Equation 8:

$$m = C \cdot M_r \cdot V \quad (8)$$

where m is the mass of copper sulfate pentahydrate crystal (in g), C is the target concentration of calibration solution (in M), M_r is the molar mass of copper sulfate pentahydrate (249.69 g/mol), and V is the volume of each calibration solution (in L).

3.3 Batch experiments

3.3.1 Short-Period Batch Experiments

To determine the optimum applied potential at different stages for copper decomposition, a series of batch experiments was conducted using solutions of varying concentrations. The process was evaluated over three consecutive reprocessing cycles. In each cycle, the wastewater concentration was held constant, and four individual batch tests were performed at one-hour intervals by applying different potential to assess which potential applied results in the highest recovery efficiency. In general, at lower concentrations, the driving force (overpotential) required to sustain the electrochemical process was higher; therefore, a more negative potential should be applied. Ideally, an inverse-relationship calibrated graph of optimum potential and waste solution concentration can be plotted, serving as a predictive tool for estimating the required optimum potential for any given concentration.

3.3.2 Long-Period Batch Experiments

A new concentration solution was prepared at the same initial value as the cycle one, as a 'real' waste solution, and the whole experiment was run for a total of 4 hours. Two different strategies were employed at this stage to optimize the electrodeposition process. Firstly, based on the short-period constructed calibration curve, process performance was implemented by adjusting the applied potential stepwise, corresponding to the real-time concentration change tested by UV-vis spectroscopy every 1 hour. Secondly, flowrate was increased from its minimum to a pump power of 2 to examine performance from a different operational perspective. The outcome of both approaches was,

consequently, compared with the outcome of applying the constant voltage to assess whether changing these parameters improves overall process performance in the aspects of FE, recovery rate and SEC.

4. Results and discussion

4.1. UV-vis Calibration Curve

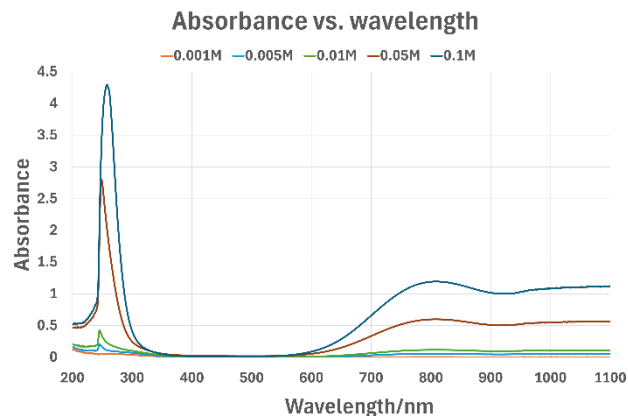


Figure 3 UV-vis spectra for five copper sulfate solutions with different concentrations, recorded from wavelength of 200 to 1100 nm, obtained from UV probe software

UV-vis spectra (Figure 3) showed two noticeable peaks: one at 280-330 nm corresponding to Cu^+-Cu^+ dimers whereas a broad peak around 750-950 nm was attributed to d-d transitions of Cu^{2+} ions. Therefore, 809 nm was selected as the characteristic wavelength for absorbance values. [13] Noteworthy, for solutions below 0.05 M, the absorption peak shifted to shorter wavelengths (hypsochromic shift or blue shift). This could be ascribed to more pronounced solvent interactions (particularly the hydrogen bonding effect of water) that alter the electronic environment of Cu^{2+} , hindering electron excitation and thus increasing the d-d energy gap. [14] When peak fell outside 890 ± 5 nm, the absorbance value at its actual peak wavelength rather than at 890 nm was recorded.

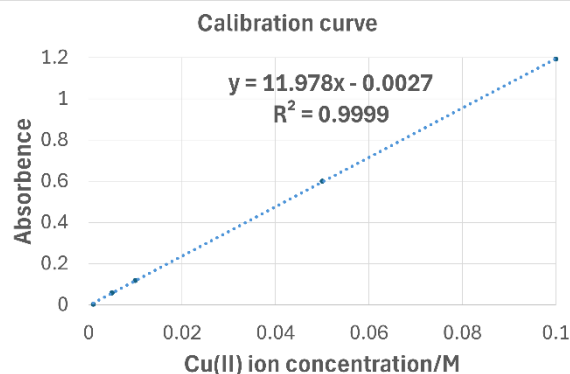


Figure 4 Calibration curve for copper concentration obtained from UV-vis spectra taking the absorbance value at a characteristic wavelength of 809 nm

As stated in Section 3.2.1, the Beer-Lambert Law predicts a linear relationship between concentration and absorbance, assuming constant molar absorptivity and unchanged cuvette thickness. A calibration curve (Figure 4) was constructed by identifying the peak absorbance of five different concentrations of copper sulfate solutions at the predetermined wavelength. Data points demonstrated a close propensity as a straight line, with a correlation factor R^2 approaching 1. This

indicated a successful calibration process, and the curve was subsequently utilized when determining the electrolyte concentrations before and after electrodeposition.

4.2. Short-Period Experiment Results

4.2.1. Current vs. Time (I-t curve)

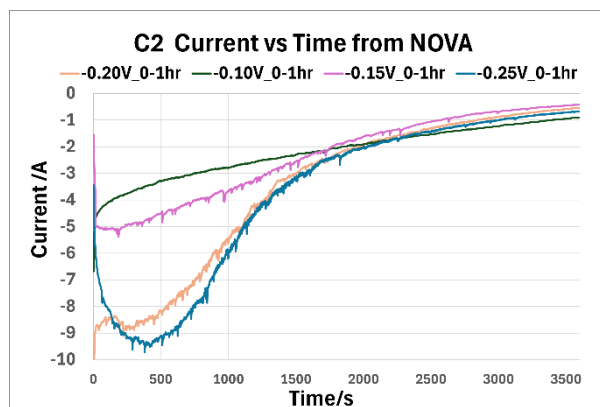


Figure 5 Current-time profiles from NOVA software for initial copper concentration=0.110 M

Figure 5 presents several typical current-time profiles from Nova, showing the performance of the copper electrodeposition in cycle two. This cycle was conducted under significant changes in concentration, which allows the study of the different kinetic regimes. As a low potential applied at -0.1 V (AgCl/Ag), the initial current started at approximately -5 A and decayed exponentially over time, and remained relatively stable at a low value, indicating the reaction was kinetic controlled by charge transfer. This was because the applied potential was only slightly negative, providing a weak driving force; hence, there was almost no concentration gradient, and the reaction rate was limited by the speed of the electron transfer process itself at the electrode surface. Increased the applied potential slightly to -0.15 V, the current started at a similar value as -0.1 V applied; however, there was a more pronounced decay throughout the reaction since it was under the mixed control region, increasing the driving force and concentration gradient slightly. This enabled some copper ions to be consumed faster at the electrode than the diffusion from the bulk solution, but the reaction was still affected by the charge transfer process initially, which means both the electrochemical reaction rate and diffusion rate are considered at these stages. Further increasing the potential applied to -0.2 V and -0.25 V, respectively, the reactions shifted to a mass transfer limited regime, particularly in -0.25 V, showing the sharpest decay curve overall. At this high potential, the initial current spike was associated with the nucleation and growth of copper deposits at the electrode and decreased rapidly to reach a plateau. This indicated the reaction rate was higher than the rate of supply of reactants, thus the system was dominated by the diffusion process. [15] [16]

In summary, to achieve optimal copper electrodeposition, an applied potential of around -0.15

V works well. This value, which is also corroborated by the calibrated curve prediction (Figure 13), lies within the mixed control regime. Consequently, operating in this region can avoid the issues of low driving force and mass limitation transfer to yield more uniform deposits and improve the overall recovery rate.

4.2.2. Current density vs. Applied potential (j-E curve)

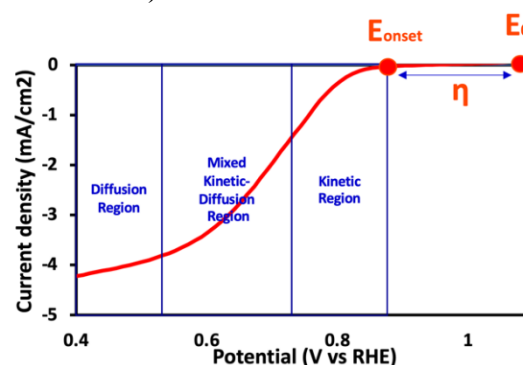


Figure 6 Conventional current density-potential (j-E) diagram showing three regions

The conventional current density-potential (j-E) diagram (Figure 6) [17] typically divides the curve into three regions, kinetic-controlled, mixed kinetic- and diffusion-controlled and diffusion-controlled region sequentially. However, the experimental j-E graph presented in Figure 7, 8 and 9 deviated from the traditional model. Specifically, an inflection point existed where the current transiently increased to a local maximum, rather than continuously decreasing toward a diffusion plateau.

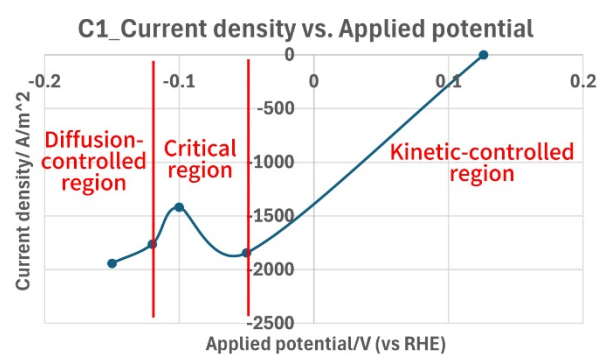


Figure 7 Experimental current density-applied potential (j-E) diagram with initial copper concentration=0.250 M

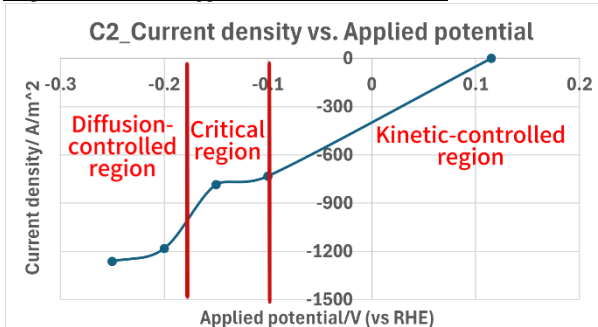


Figure 8 Experimental current density-applied potential (j-E) diagram with initial copper concentration=0.110 M

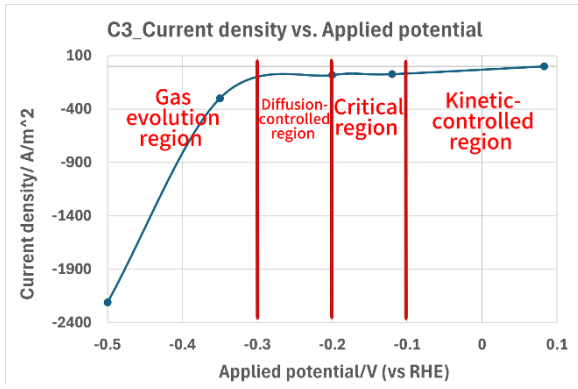


Figure 9 Experimental current density-applied potential (j - E) diagram with initial copper concentration = 0.090 M

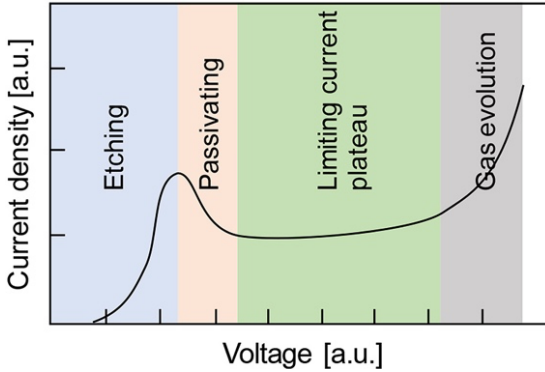


Figure 10 Current density-potential diagram from Jacquet's study about metal dissolution demonstrating four regions

Jacquet's four-regime model (etching, passivating, limiting-current plateau and gas-evolution) shown in Figure 10, provides an instrumental framework for interpreting this anomaly. [18] Although Jacquet studied anodic metal dissolution rather than cathodic metal deposition, the j - E diagram demonstrated a remarkable similarity in overall shape. By integrating inspiration from Jacquet's four-regime approach with traditional three-region theory, the experimental j - E relationship could be classified into four regions.

Region 1: Kinetic-controlled region

At low overpotentials, mass transport was sufficiently fast so that copper concentration at the electrode surface was almost equivalent to bulk concentration. The electrodeposition rate was governed by reaction kinetics; an increasing overpotential led to a more negative current density.

Region 2: Critical region (Mixed kinetic- and diffusion-controlled region)

A temporary rise in current was observed at intermediate overpotentials. As reaction rate increased, partial pore blockage of the nickel foam electrode and the formation of a hydrodynamic boundary layer [8] (due to the inconsistency of electrolyte velocities) hindered the electrochemical reaction. This resembled the passivating zone proposed in Jacquet's model, though the underlying mechanism differs. With further increase in overpotential, the system conquered this restriction, and the current became more negative again.

Region 3: Diffusion-controlled region

At high overpotentials, the copper ions were being reduced instantaneously upon reaching the electrode surface. There existed a steep concentration gradient

between the electrode surface and the bulk electrolyte. The system eventually reached a diffusion-limited plateau.

Region 4: Gas evolution region

Hydrogen evolution as the side reaction at the cathode became dominant at extremely high overpotentials, leading to a noticeable slump in current. This effect was most apparent at low starting concentrations. For higher concentrations, equipment constraints (maximum current of 10 A) prevented reaching this region.

Figures 7, 8 and 9 also disclosed the link between initial slopes of the j - E curve with solution concentrations. This discovery could be explained using the Butler-Volmer equation. The completed form is given in Equation 9 below. [19]

$$j = j_0 \times \left\{ \frac{c_o(0, t)}{c_o^*} \exp \left[\frac{\alpha_a z F \eta}{RT} \right] - \frac{c_r(0, t)}{c_r^*} \exp \left[\frac{\alpha_c z F \eta}{RT} \right] \right\} \quad (9)$$

Focusing on the kinetic-controlled region with the assumption of negligible concentration gradient between electrode surface and bulk liquid, the simplified Butler-Volmer equation for cathodic reaction could be written as:

$$j_c = -j_0 \times \exp[-\alpha_c f \eta] \quad (10)$$

where j_c is the cathodic current density, j_0 is the exchange current density, α_c is the cathodic charge transfer coefficient, f is a constant ($f = \frac{zF}{RT}$), η is the overpotential. [20] For the same overpotential and assuming a constant α_c for all experiments, the cathodic current density depended solely on the exchange current density, which itself was concentration dependent. [20] As a result, a higher initial copper concentration led to a larger j_0 , and thus a steeper initial slope in the j - E curve and could reach a more negative cathodic current. Additionally, for the lowest initial concentration (0.009 M), the onset potential was most conspicuous, whereas the onset potential for higher initial concentrations (0.110 M and 0.250 M) was nearly indistinguishable from the equilibrium potential.

4.2.3. Faradaic efficiency vs. Applied potential (FE-E curve)

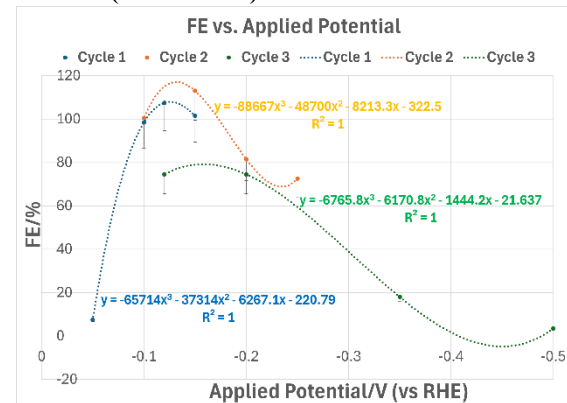


Figure 11 Faradaic efficiency against applied potential (FE-E) diagram for three sets of experiments with different initial concentrations (cycle 1 = 0.250 M, cycle 2 = 0.110 M, cycle 3 = 0.009 M) approximated using a third-order polynomial

The Faradaic Efficiency (FE) was obtained via two approaches: mass change of electrode and the concentration change of the electrolyte. However, some data points yielded FE values surpassing 100%, which are physically unrealistic. Mass-based overestimation could be attributed to water trapped in the nickel foam and oxidation of deposited copper during drying. Concentration-based FE was overestimated due to imperfect mixing of electrolyte and volume loss during repeated sampling.

Averaged FE values were plotted against the applied potential for each experimental cycle, as shown in Figure 11. The resulting curves exhibited a non-linear, asymmetric trend with a salient maximum point, which is consistent with literature about carbon dioxide reduction systems. [21] A cubic equation was employed to approximate the *FE-E* behavior with correlation factors approaching 1. However, lower order polynomials illustrated poor alignment with the experimental data, whereas higher-order functions tended to overfit the data and potentially introducing unnecessary complexities. The *FE-E* trend aligned with the *j-E* interpretations. At low overpotential, copper reduction was insufficient and thus led to low FE. Increasing overpotential enhanced copper reduction with minimal side reaction, giving a rise in FE. A maximum point was observed within the critical region where the reaction was neither kinetic nor diffusion limited. At high overpotentials, constricted Cu^{2+} mass transport and spike in hydrogen evolution shifted the current selectivity away from copper reduction, resulting in reduced FE. The optimum potential was calculated by applying the first-order derivative for the best-fit curve functions, allowing all data to be incorporated rather than relying on discrete points.

4.2.4. Specific energy consumption vs. Applied potential (SEC-E curve)

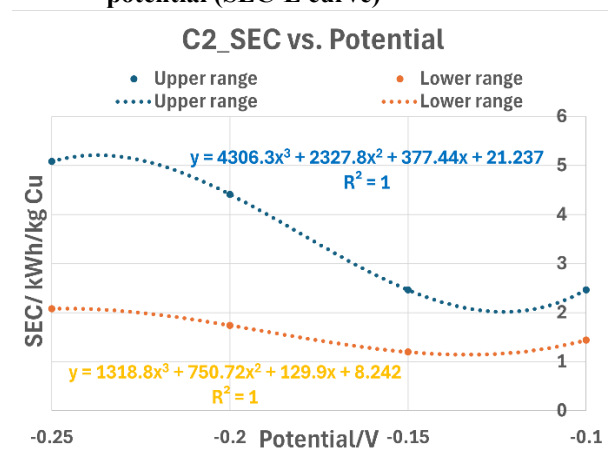


Figure 12 Specific energy consumption against applied potential (SEC-E) diagram for initial copper concentration=0.110 M

The specific energy consumptions (SEC) were determined and presented as a range for each data point. Similar to the FE-E analysis, the SEC-E relationship was also modelled using cubic equations, as shown in Figure 12 below. SEC and FE exhibited an inverse correlation as illustrated in Equation 6, where higher FE generally led to lower SEC.

Consequently, the minimum SEC was expected to lie near the region where FE reached its maximum.

The optimal potential with respect to SEC was obtained by differentiating the cubic function. Notably, the optimum potential derived from the FE-E and SEC-E curves did not coincide exactly. This was reasonable because FE and SEC reflect different aspects of sustainability. Hence, the observed discrepancy emphasized the need to balance the current selectivity and energy efficiency in practical electrodeposition operations.

4.2.5. Optimum applied potential vs. Concentration (E*-C curve)

Corresponding to the previous FE and SEC optimum points identified, a calibrated curve correlating the optimum applied potential with electrolyte concentration was conducted, representing Figure 13. It was evident to see a clear downward decay shape. As concentration increases the optimum potential becomes less negative gradually, which means a more negative potential is required to drive the deposition reaction at a sufficient rate when concentration becomes lower. This inverse relationship indicated that in dilute solution, the need of overpotential not only made the deposition process of copper more challenging but also suggested the higher operational cost, such as the capital cost input for the power supply system.

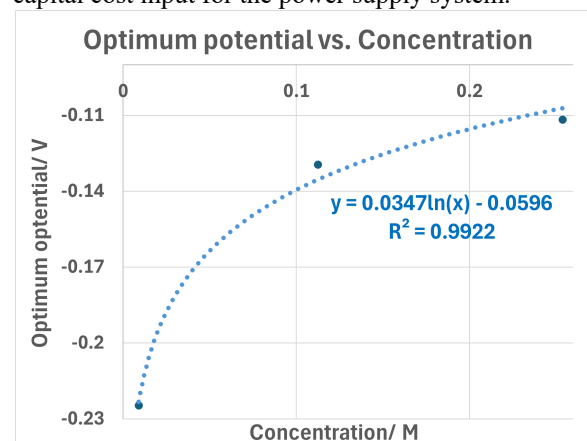


Figure 13 Optimum applied potential against electrolyte concentration (E*-C) diagram

Moreover, to be consistent, this curve was optimized by the logarithmic function, which mirrored the form of the fundamental Nernst equation applied for calculating the equilibrium potential. Each data point on this curve represented an optimized pair of applied potential and specific concentration that simultaneously maximized FE and minimized SEC. Based on this, an appropriate applied potential was selected and utilized for the long-term experiment in each hour.

4.2.6. Equilibrium potential vs. Concentration (E-C curve)

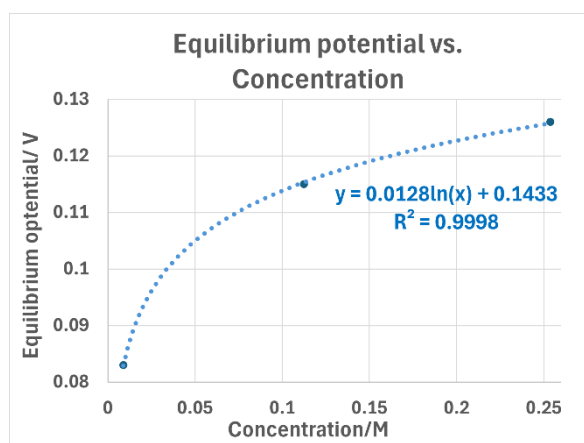


Figure 14 Equilibrium potential relative to the reference electrode and electrolyte concentration (E-C) diagram

Figure 14 illustrates the theoretical relationship between the equilibrium potential relative to the reference electrode (RHE) and electrolyte concentration, as governed by the concept of Nernst equation. As shown in this graph, the concentration gradient is the key factor that affects the magnitude of equilibrium, confirming that the larger the concentration differences, the larger the equilibrium potential required to balance it. Mathematically, it is obvious that the slope constant of this line is identical with the theoretical Nernst slope at 298 K, $RT/2F=0.00128$ for two-electrons transfer, aligning perfectly with the Nernst equation of copper reduction, and the y-intercept is the offset from the reference electrode. As the same trend was shown with the optimum potential and concentration graph previously, this sets the thermodynamic baseline for the predicted optimum potential and investigate overpotential and reaction kinetics later.

4.3. Long-Period Experiment Results

4.3.1. Recovery vs. time

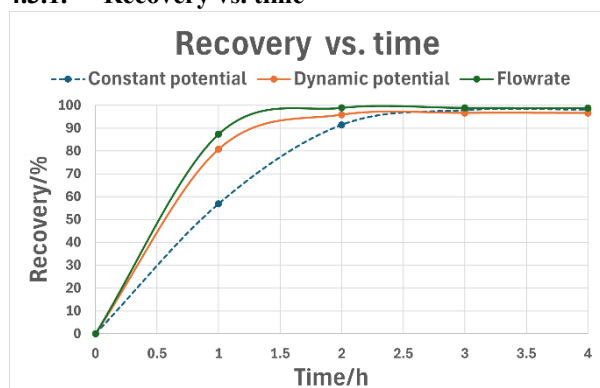


Figure 15 Recovery against time diagram for three operational modes in long-period experiments

The recovery percentage for 3 operational modes was compared over time, as shown in Figure 15. As expected, the higher flow rate achieved the highest recovery at around 95% most rapidly, followed by a stable plateau, with dynamic potential and constant potential performed sequentially slower. A high flowrate thinned the diffusion layer in electrodeposition by creating greater turbulence, which enabled reacted species to be removed from the electrode frequently and the bulk solution to flow rapidly across the electrode surface, prevented the

formation of a thick blocking layer. [22] [23] With sustained high current in the initial period, a high flow rate resulted in the steepest slope on the recovery graph. In contrast, constant potential yielded the slowest slope as it quickly became mass-transfer limited, allowing diffusion layer to build up and hence reducing the overall disposition rate. Dynamic potential is represented as an intermediate case. Its recovery is faster than constant potential because the optimal potential was adjusted to always push the surface reaction as fast as possible, though it was less effective at mitigating mass-transfer limitations than a high flow rate strategy.

Additionally, the three curves eventually converged due to low bulk concentration and electrode passivation, such as formation of dendrites and a porous layer. [24] These non-uniform coatings might physically block the underlying electrode surface, preventing new ions from reaching the active sites. [25] Consequently, the experiment hit the ceiling as the deposition rate approached zero, and the recovery percentage could no longer increase further.

4.3.2. FE vs. time and pH vs. time

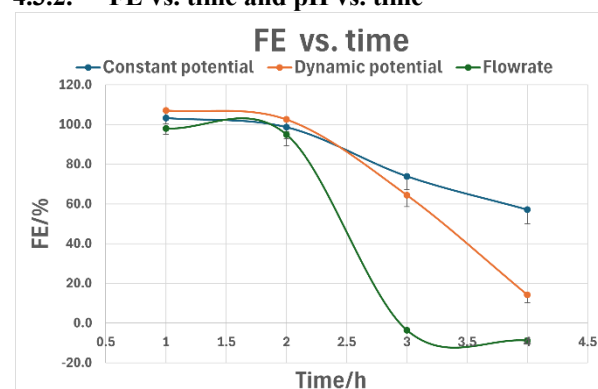


Figure 16 Faradic Efficiency against time diagram for three operational modes in long-period experiments

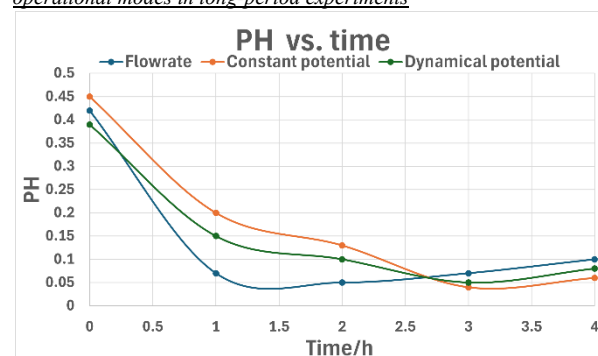


Figure 17 pH against time diagram for three operational modes in long-period experiments

Another key insight comes from examining the relationship between FE and time (Figure 16). There is always a trade-off between recovery and FE in the electrodeposition process: the balance between the rate of copper deposition (recovery) and the selectivity of electron utilization (FE). Therefore, the performance of the three operational modes in these two aspects was inversely correlated, as clearly shown in the two corresponding graphs. Initially, all three types of experiments operated at a similarly high FE. However, as the reaction progressed, their performance diverged

significantly. Constant potential maintained the highest and most stable FE. In contrast, increased flow rate illustrated the most severe drop in FE. The main reason was the activation of competing side reactions since the electrical current was sacrificed to produce hydrogen gas. [26]

A pH-versus-time graph (Figure 17) provided evidence for this phenomenon. All three curves showed an initial decrease in pH, which can be attributed to water oxidation at the anode. [27] Subsequently, the observed slight increase in PH was consistent with the hydrogen bubbles during HER. These bubbles increased the mass transfer resistance, hindered metal ion transfer for further deposition. [28] [29] The condition with an increased flow rate showed the greatest pH variation. This was because a high flow rate promotes rapid deposition, thus more intense local ion depletion. With fewer copper ions available, the same applied potential now preferentially drove HER, produced more hydrogen bubbles and emphasized the relation between pH and efficiency.

4.3.3. Economic analysis

A cost comparison was carried out to assess the economic viability of copper electrodeposition under various modes. Since the recovery process approached a plateau after two hours, the cost for the first two hours of electrodeposition was examined and compared with the cost of purchasing freshly mined copper [30]. The economic analysis focused on the expenses related to the applied potential and pumping only, excluded all other costs such as electrode materials, labor and maintenance. The cost of the applied potential was calculated using a range of SEC values, while the pump power consumption was taken at between 29 and 34 W [31]. The cost for electricity was taken as 0.17 £/kWh. [32]

Method	Lower cost (£/tonne of Cu)	Upper cost (£/tonne of Cu)	Average cost (£/tonne of Cu)
Constant potential with minimum flowrate	555	757	656
Dynamic potential with minimum flowrate	499	666	583
Constant potential with higher flowrate	531	983	757
Freshly mined copper	-	-	8392

Table 1 Economic analysis comparing the cost of three electrochemical operational strategies with the cost of freshly mined copper

Evidently, the average market cost of freshly mined copper is approximately an order of magnitude higher than the cost for electrodeposition. This highlighted the economic advantage of using electrodeposition for copper recovery. Furthermore, dynamic potential operation at the minimum flow rate exhibited the lowest average cost among the three electrochemical methods. This indicated that dynamic potential control

can simultaneously reduce energy consumption and enhance recovery efficiency.

5. Conclusion

This study demonstrated the relationship between applied potential and process performance (FE and SEC). An optimal applied potential was identified for each initial waste solution concentration. This correlation was described by a logarithmic equation, analogy to Nernst equation. In long period tests, dynamically adjusting the applied potential revealed the lowest operational cost (£ 583 per ton of Cu) due to optimized current efficiency and energy consumption. Additionally, utilizing higher flowrate with constant applied potential reached the highest copper recovery of 98.9 % due to enhanced mass transfer.

Overall, the investigation showed a possibility of integrating dynamic potential control with optimized flow rate to make the process more efficient and cost-effective.

6. Outlook

This study on copper electrodeposition could be improved in several aspects: experimental accuracy, process limitation, and sustainability.

6.1. Accuracy improvement

Future experiments should implement more rigorous procedures to minimize measurement errors. To address the inaccuracy caused by water trapped inside the nickel foam electrode, extending the drying duration with a drying oven is recommended. Also, the mass of the electrode could be recorded at a fixed time interval (e.g., every 15 minutes) to construct a mass change curve with respect to time. By identifying the mass plateau to obtain a more accurate electrode mass change. The overestimation of bulk concentration due to imperfect mixing could be ameliorated by introducing a continuous stirring set up. Specifically, the reservoir could be put onto a magnetic stirrer and sampling several minutes after the reaction has been stopped to facilitate mixing and a more congruent concentration distribution in the reservoir.

6.2. Process optimization

Although the current study reveals that the electrodeposition process is effective in reducing copper concentration by several order of magnitudes, the copper (II) ion concentration reached a plateau around 100-200 mg/L after two hours of electrodeposition, which unsatisfied with the dischargeable standard proclaimed by the Environmental Agency (1 mg/L). This indicates the necessity of further electrodeposition optimization or hybrid treatment strategies before the copper solution can be disposed directly into river. Firstly, an increased flowrate at constant applied potential obtained the highest copper recovery among the three long-period experiments, reveals the optimization possibility of flowrate. Theoretically, a higher flowrate could motivate mass transport, retrench the concentration gradient and thus allow more complete deposition of

copper ions. Future work should systematically explore higher flowrates, while balancing the gains in copper recovery against associated pumping cost and potential undesirable turbulence. Secondly, performance of cells with different configurations such as bipolar electrodes, continuous stirred tank and alternating current, could be examined. [33] Thirdly, given that electrodeposition along may not achieve the final goal for regulatory compliance. Electrodeposition could be served as a primary removal technique followed by a secondary treatment such as absorption or chemical precipitation, which are particularly effective in low concentration regimes. [34][35]

6.3. Environmental evaluation

A further opportunity for improving the sustainability of the electrodeposition process is to look for alternatives for the cathode material. The present work utilized nickel foam as the working electrode due to its excellent electrical conductivity and enormous surface area. However, nickel is regarded as a carcinogenic substance and nickel mining is associated with significant ecological impacts. [36][37] Nickel is more reactive than copper and thus may dissolve into the electrolyte during electrodeposition, further contaminate the waste solution. Graphite is inert, cost-effective and more environmentally friendly than nickel, offering better stability and recyclability. The high affinity of copper electrode to copper (II) ions could facilitate the nucleation of copper onto electrode. This may lead to a more uniform copper deposition. Furthermore, a copper electrode could further simplify the recovery process. In contrast, further special treatment, such as acid wash, is required for graphite and nickel electrodes. Although graphite poses extra environmental issues, a comprehensive life cycle analysis (LCA) comparing these electrode materials could guide more sustainable choices in future applications.

7. Reference

[1] Biswas, S. and Mishra, U. (2016). Treatment of Copper Contaminated Municipal Wastewater by Using UASB Reactor and Sand-Chemically Carbonized Rubber Wood Sawdust Column. *BioMed Research International*, [online] 2016, pp.1–9. doi: <https://doi.org/10.1155/2016/5762781>.
 [2] Subramanya, M. (n.d.). *Streams*. [online] Available at: <https://www.ijert.org/research/electrowinning-of-copper-from-industrial-waste-streams-IJERTV10IS080071.pdf>.
 [3] Smriti, Ahmed, A., Lodhi, S. and Shukla, S. (2023). Copper toxicity in aquatic ecosystem: A Review. *International Journal of Fisheries and Aquatic Studies*, 11(4), pp.134–138. doi: <https://doi.org/10.22271/fish.2023.v11.i4b.2835>.
 [4] Bost, M., Houdart, S., Oberli, M., Kalonji, E., Huneau, J.-F. and Margaritis, I. (2016). Dietary copper and human health: Current evidence and unresolved issues. *Journal of Trace Elements in Medicine and Biology*, [online] 35, pp.107–115. doi: <https://doi.org/10.1016/j.jtemb.2016.02.006>.

[5] Podobińska-Staniec, M., Wiktor-Sułkowska, A., Kustra, A. and Sylwia Lorenc-Szot (2025). Copper as a Critical Resource in Energy Transition. *Energies*, [online] 18(4), pp.969–969. doi: <https://doi.org/10.3390/en18040969>.
 [6] Li, S., Dai, M., Ali, I., Bian, H. and Peng, C. (2023). A new idea for efficient copper recovery from wastewater by electrodeposition: Adsorption pretreatment. *Desalination*, 562, pp.116683–116683. doi: <https://doi.org/10.1016/j.desal.2023.116683>.
 [7] Huang, X., Shi, X. and Zeng, H. (2023). How far does the Copper/Nickle recover from the practical application in the electroplating wastewater? *Resources, conservation & recycling advances*, 19, pp.200170–200170. doi: <https://doi.org/10.1016/j.rcradv.2023.200170>.
 [8] Esmacili, H. and Foroutan, R. (2015). INVESTIGATION INTO ION EXCHANGE AND ADSORPTION METHODS FOR REMOVING HEAVY METALS FROM AQUEOUS SOLUTIONS. [online] 4(12), pp.620–629. Available at: https://www.researchgate.net/publication/289673360_INVESTIGATION_INTO_ION_EXCHANGE_AND_ADSORPTION_METHODS_FOR_REMOVING_HEAVY_METALS_FROM_AQUEOUS_SOLUTIONS.
 [9] From supervisor Anna Hankin ‘Copper electrodeposition study’ handout
 [10] Zhu, J., Chen, B., Ni, M. and Xu, H. (2024). A mini-review on three-electrode configuration for Solid Oxide Cells. *Energy Reviews*, pp.100128–100128. doi: <https://doi.org/10.1016/j.enrev.2024.100128>.
 [11] <https://www.facebook.com/IEST.XM> (2024). *The Three Electrode System: Key to Electrochemical Research*. [online] IEST | Lithium Battery Tester | One-Stop Li-ion Battery Testing Solution Provider. Available at: <https://iestbattery.com/case/introduce-the-three-electrode-system/>.
 [12] Rocha, F.S., Gomes, A.J., Lunardi, C.N., Kaliaguine, S. and Patience, G.S. (2018). Experimental methods in chemical engineering: Ultraviolet visible spectroscopy-UV-Vis. *The Canadian Journal of Chemical Engineering*, 96(12), pp.2512–2517. doi: <https://doi.org/10.1002/cjce.23344>.
 [13] Hung, S.-C., Lu, C.-C. and Wu, Y.-T. (2021). An Investigation on Design and Characterization of a Highly Selective LED Optical Sensor for Copper Ions in Aqueous Solutions. *Sensors*, [online] 21(4), pp.1099–1099. doi: <https://doi.org/10.3390/s21041099>.
 [14] Robinson, J.W., Skelly, E.M. and Frame, G.M. (2004). *Undergraduate Instrumental Analysis, Sixth Edition*. CRC Press.
 [15] Timmerhuis, N.A.B., Wood, J.A. and Lammertink, R.G.H. (2021). Connecting experimental degradation kinetics to theoretical models for photocatalytic reactors: The influence of mass transport limitations. *Chemical Engineering Science*, 245, pp.116835–116835. doi: <https://doi.org/10.1016/j.ces.2021.116835>.
 [16] D’Ajello, P.C.T. and Schervenski, A.Q. (2004). Current-time and current-potential profiles in electrochemical film production. (II) Theoretical current-potential curves. *Journal of Electroanalytical Chemistry*, 573(1), pp.37–42. doi: <https://doi.org/10.1016/j.jelechem.2004.06.019>.

- [17] Conventional j-E diagram from Sustainable lecture slides
- [18] Han, W. and Fang, F. (2019). Fundamental aspects and recent developments in electropolishing. *International Journal of Machine Tools and Manufacture*, 139, pp.1–23. doi: <https://doi.org/10.1016/j.ijmachtools.2019.01.001>.
- [19] Bard, A.J. and Faulkner, L.R. (2001). *Electrochemical methods fundamentals and applications*. New York: John Wiley & Sons.
- [20] Dickinson, E.J.F. and Wain, A.J. (2020). The Butler-Volmer equation in electrochemical theory: Origins, value, and practical application. *Journal of Electroanalytical Chemistry*, 872, p.114145. doi: <https://doi.org/10.1016/j.jelechem.2020.114145>.
- [21] Feroze, M.T., Sami, S.K., Doonyapisut, D., Kim, B. and Chung, C. (2020). Electrochemical Reduction of CO₂ into C1 and C2 Hydrocarbons Using Dendritic Cu and Cu₂O Electrodes. *ChemElectroChem*, 7(3), pp.730–736. doi: <https://doi.org/10.1002/celec.201902035>.
- [22] Klement, N., Savino, E., Rooijmans, S., Patty, Lynn, N.S., Browne, W.R. and Verpoorte, E. (2025). Electrochemical Flow Reactors: Mass Transport, iR Drop, and Membrane-Free Performance with In-Line Analysis. *ACS electrochemistry*. doi: <https://doi.org/10.1021/acselectrochem.4c00167>.
- [23] Zhang, H., Zhu, J., Wang, C., Yuan, H., Dai, H. and Wei, X. (2025). Experimental Study on the Impact of Flow Rate Strategies on the Mass Transfer Impedance of PEM Electrolyzers. *Energies*, [online] 18(11), pp.2700–2700. doi: <https://doi.org/10.3390/en18112700>.
- [24] Mistry, A. and Srinivasan, V. (2019). On our Limited Understanding of Electrodeposition. *MRS Advances*, 4(51-52), pp.2843–2861. doi: <https://doi.org/10.1557/adv.2019.443>.
- [25] Kumar, S., Pande, S. and Verma, P. (2015). Factor Effecting Electro-Deposition Process. 700| *International Journal of Current Engineering and Technology*, [online] 5(2). Available at: <https://inpressco.com/wp-content/uploads/2015/03/Paper18700-703.pdf>.
- [26] Alebrahim, M.F., Khattab, I.A. and Sharif, A.O. (2015). Electrodeposition of copper from a copper sulfate solution using a packed-bed continuous-recirculation flow reactor at high applied electric current. *Egyptian Journal of Petroleum*, [online] 24(3), pp.325–331. doi: <https://doi.org/10.1016/j.ejpe.2015.07.009>.
- [27] Alebrahim, M.F., Khattab, I.A. and Sharif, A.O. (2015). Electrodeposition of copper from a sulfate copper using a packed-bed continuous-recirculation flow reactor at high applied electric current. *Egyptian Journal of Petroleum*, [online] 24(3), pp.325–331. doi: <https://doi.org/10.1016/j.ejpe.2015.07.009>.
- [28] Turmiasaputri, A.I.N., Erfan Handoko and Budi, S. (2022). The Effect of Temperature and pH on FeCoNi Film Electrodeposition. *Chemistry and Materials*, 1(2), pp.61–65. doi: <https://doi.org/10.56425/cma.v1i2.25>.
- [29] Nikitin, V.S., Ostanina, T.N., Rudoi, V.M., Kuloshvili, T.S. and Darintseva, A.B. (2020). Features of hydrogen evolution during electrodeposition of loose deposits of copper, nickel and zinc. *Journal of Electroanalytical Chemistry*, [online] 870, p.114230. doi: <https://doi.org/10.1016/j.jelechem.2020.114230>.
- [30] The (2025). *LME Copper | London Metal Exchange*. [online] Lme. Available at: <https://www.lme.com/Metals/Non-ferrous/LME-Copper#Overview>.
- [31] KNF Group. (2025). *KNF Group*. [online] Available at: <https://knf.com/en/us> [Accessed 10 Dec. 2025].
- [32] About this release. (n.d.). Available at: <https://assets.publishing.service.gov.uk/media/68da5e91dadf7616351e4b5e/quarterly-energy-prices-september-2025.pdf>.
- [33] Al-Qodah, Z. and Al-Shannag, M. (2017). Heavy metal ions removal from wastewater using electrocoagulation processes: A comprehensive review. *Separation Science and Technology*, pp.1–28. doi: <https://doi.org/10.1080/01496395.2017.1373677>.
- [34] Liu, Y., Wang, H., Cui, Y. and Chen, N. (2023). Removal of Copper Ions from Wastewater: A Review. *International journal of environmental research and public health*, [online] 20(5). doi: <https://doi.org/10.3390/ijerph20053885>.
- [35] Gong, Y., Wang, Y., Tang, M., Zhang, H., Wu, P., Liu, C., He, J. and Jiang, W. (2022). A two-step process coupling photocatalysis with adsorption to treat tetracycline - Copper (II) hybrid wastewaters. *Journal of Water Process Engineering*, [online] 47, p.102710. doi: <https://doi.org/10.1016/j.jwpe.2022.102710>.
- [36] UK (2025). *Nickel: toxicological overview*. [online] GOV.UK. Available at: <https://www.gov.uk/government/publications/nickel-properties-incident-management-and-toxicology/nickel-toxicological-overview>.
- [37] Lo, M.G.Y., Morgans, C.L., Santika, T., Mumbunan, S., Winarni, N., Supriatna, J., Voigt, M., Davies, Z.G. and Struebig, M.J. (2024). Nickel mining reduced forest cover in Indonesia but had mixed outcomes for well-being. *One Earth*, [online] 7(11), pp.2019–2033. doi: <https://doi.org/10.1016/j.oneear.2024.10.010>.

Process Design of a Hydrogen Liquefaction Plant Operated with Intermittent Wind Electrolysis

Hakan Koçan, Winston Tang

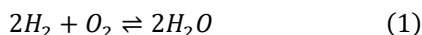
Department of Chemical Engineering, Imperial College London, U.K.

Abstract: Hydrogen is emerging as a promising energy carrier, yet its potential is constrained by the challenges in storage and transport. Liquid hydrogen addresses these issues, as it offers greater volumetric energy density. Although liquid hydrogen is projected to have an increase in demand as a carrier, nearly all plants rely on fossil-derived hydrogen, and no established designs support fully renewable production. Renewably produced hydrogen suffers from high intermittency, making steady-state operation impossible. This study develops a flexible hydrogen liquefaction system where hydrogen gas is supplied by electrolysis powered by wind turbines, integrating a dynamic storage buffer and a throughput-adaptable liquefaction process. A model was adopted and developed in Aspen Plus, rating fixed-area heat exchangers and simulating ortho-para hydrogen conversion, with a reliquefaction sub-train for flash gas recovery. To manage the variability in hydrogen production, a storage tank was modelled on MATLAB, and a feedback controller was designed to smooth the fluctuating inlet flows. Multiple liquefaction configurations were evaluated for energy performance across varying throughputs. The final design was a two-train configuration operating across 1-60 kg/hr, achieving an average specific energy consumption (SEC) of 16.47 kWh/kg_{LH2} under optimal operating conditions. The findings show that hydrogen gas buffering effectively overcomes the renewable intermittency and allows stable liquefaction, whilst the multi-train configuration reduces the SEC throughout its full range. Although certain limitations still remain in tank modelling accuracy, ramp-up complexity, and liquid nitrogen energy consumption, the study provides one of the first conceptual designs for renewably produced hydrogen liquefaction.

1. Introduction

In 2022, global CO₂ emissions reached an all-time high of 36.1 Gt, with the power sector responsible for almost 40% of this total (Liu et al., 2023). Despite continued efforts to reduce fossil energy use, global carbon emissions continue to increase steadily, as shown in Figure 1. This highlights the limitations of current decarbonisation strategies and the urgency of transitioning to renewable energy.

Recently, hydrogen has emerged as an attractive alternative to fossil fuels. Unlike traditional fossil fuels, which release a significant amount of carbon dioxide when combusted, hydrogen only produces water vapour when used as fuel (Koons, 2023).



Hydrogen can also be produced via electrolysis using renewable energy sources such as solar and wind power, which have a near-zero carbon footprint. While hydrogen has the disadvantage of a low volumetric energy density, the liquefaction of hydrogen addresses this issue by significantly reducing its volume, allowing convenient

transport and storage of hydrogen in lorries or trailers. The potential of hydrogen as an energy carrier is expected to increase global hydrogen demand significantly. By 2050, the total hydrogen demand is projected to reach up to 234 Mt, with liquid hydrogen accounting for roughly 17% of this amount (Alanazi et al., 2025).

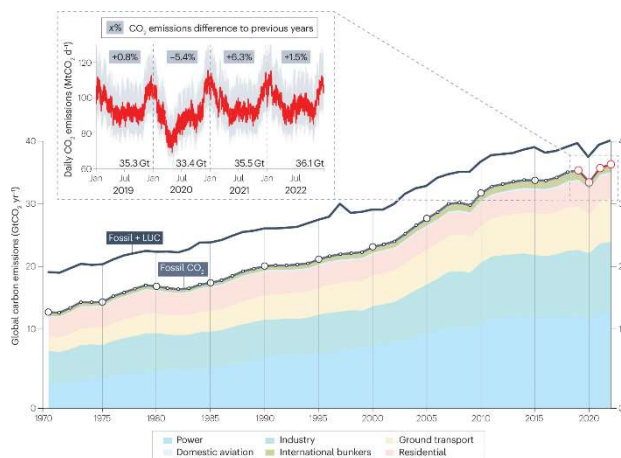


Figure 1: Global CO₂ emissions from 1970 to 2022. (Liu et al., 2023)

Despite this, the development of renewable hydrogen is far from complete. Nearly all liquefaction plants in the last two decades have utilised steam reforming as their source of hydrogen. One major issue with renewable hydrogen is the difficulty in maintaining a stable supply. The power generation of renewable sources often depends on local weather conditions, which fluctuate significantly over time, causing the production rate of hydrogen gas to vary accordingly.

Although hydrogen liquefaction is a well-established process in literature, current research continues to target the reduction in specific energy consumption of steady-state processes. For example, the conceptual liquefaction process developed by IDEALHY focuses on improving process efficiency, while assuming a steady, constant feed. (Essler et al., 2012) Much less work has examined how liquefaction plants should be designed when subjected to unstable hydrogen throughput.

The objective of this study is to address the gap in research by developing a flexible liquefaction process designed to accommodate a wide range of hydrogen flow rates. Additionally, this study aims to design a buffer control strategy that mitigates the unstable hydrogen production rate of wind turbines.

2. Background

2.1 Liquefaction in Industry

At present, Linde's Leuna liquefaction plant remains the only liquefaction facility capable of utilising renewably produced hydrogen. Despite having the world's largest PEM electrolyser, renewably produced hydrogen contributes only 5% of the plant's hydrogen supply, while more than 90% of the liquefier's hydrogen is derived from fossil fuels. (Linde-engineering.com, 2021)

2.2 Liquefaction Fundamentals

While industrial hydrogen liquefaction plants often differ in production capacity, cooling fluids, exchanger networks, and final expansion methods, they all follow a generic liquefaction process. First, hydrogen gas is compressed and cooled to an intermediate temperature. It is then further cooled to temperatures near its boiling point through a network of heat exchangers, while *ortho*-hydrogen is also catalytically converted to *para*-hydrogen during this step. Finally, the fluid is expanded and partially

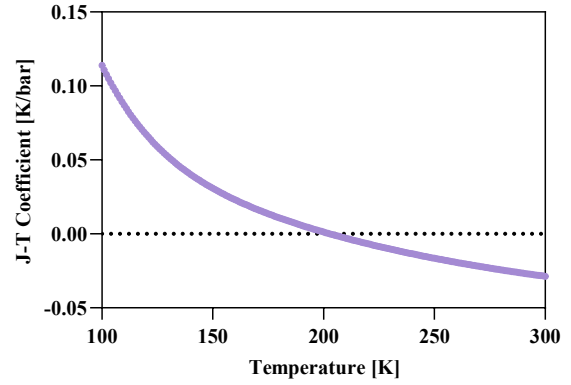


Figure 2: Hydrogen J-T coefficient with temperature

liquefied, and the liquid hydrogen is separated from the mixture and stored.

2.2.1 Joule-Thomson Coefficient and Precooling

The final step of a hydrogen liquefaction process often involves a Joule-Thomson (J-T) valve, where hydrogen is cooled via adiabatic expansion, producing a vapour-liquid mixture. This phenomenon is referred to as the Joule-Thomson effect. The following expression describes the pressure-temperature relationship under adiabatic conditions:

$$\mu_{JT} = \left(\frac{\partial T}{\partial P} \right)_H \quad (2)$$

Where μ_{JT} is the J-T coefficient. At ambient temperatures, μ_{JT} of hydrogen is negative, i.e. a drop in pressure results in an increase in temperature. μ_{JT} becomes positive only when hydrogen is cooled below its inversion temperature, which is documented to be around 200 K at atmospheric pressure (NIST, 2023). Therefore, hydrogen must be pre-cooled below its inversion temperature before entering the J-T valve. Otherwise, throttling cannot produce the cooling required for liquefaction.

2.2.2 Ortho- to Para- Hydrogen Conversion

Molecular hydrogen occurs in two nuclear spin isomers, namely *ortho*- and *para*-hydrogen. In *ortho*-hydrogen, the spins of the atoms are parallel, giving rise to 3 degenerate spin states. On the other hand, *para*-hydrogen has anti-parallel spins and a single spin state. Therefore, at high temperatures, where many rotational energy levels are accessible, *ortho*-hydrogen becomes more prevalent due to its greater degeneracy. Statistical mechanics show that the equilibrium mixture at ambient temperature contains 75% *ortho*- and 25% *para*-hydrogen (Maczek and Meijer,

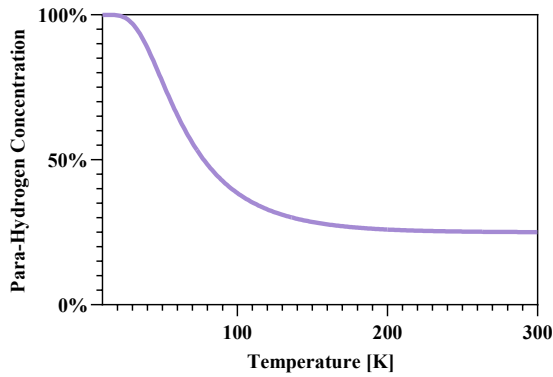


Figure 3: Para-Hydrogen concentration as a function of

2017). However, as the temperature decreases, only low rotational energy levels become accessible. As *para*-hydrogen corresponds to the lowest energy state, the equilibrium shifts heavily towards *para*-hydrogen, and almost all hydrogen exists in *para*-hydrogen form at cryogenic temperatures.

The *ortho*- to *para*- conversion of hydrogen is a slow, exothermic process. Therefore, if this conversion occurs during storage, it could lead to an increase in temperature, causing excess boil-off gas accumulation in the tank, which poses serious explosion risks. To prevent this, *ortho*-hydrogen is often catalytically converted to *para*-hydrogen during the liquefaction process.

3. Methodology

3.1 Base Liquefaction Model

An initial steady-state process design was adopted from Yin and Ju (2020), as it was optimised for small-scale production, and the operation units, flow rates, and parameters were clearly stated, making it a suitable starting point for modelling.

First, hydrogen is pre-cooled using liquid nitrogen in two multi-stream heat exchangers (MHEXs). The chilled hydrogen is then further cooled in a series of 6 heat exchangers, with helium acting as the coolant. The helium is compressed and cooled before being reintroduced to the heat exchanger network as a hot stream. Between the heat exchanger sections, *ortho*- to *para*- conversion of hydrogen is carried out in catalytic reactors. Finally, the hydrogen is throttled through a J-T valve, where the pressure decreases from 24 to 1.5 bars, producing a mixture of gaseous and liquid hydrogen.

3.2 Assumptions

Multiple assumptions were made to simplify the model:

- All impurities were removed from the hydrogen stream before entering the liquefaction process.
- No pressure drop across heat exchangers, coolers, and heaters.
- Isentropic efficiencies of expanders and compressors are set to 0.8.
- Overall heat transfer coefficient U is constant with a value of $0.85 \text{ kWm}^{-2}\text{K}^{-1}$.

3.3 Equation of State Selection

Valenti et al. showed that the choice of equation of state and heat-capacity correlations can result in differences in calculated exergy of up to 13%. Although Leachman et al. have already developed a reference EOS for hydrogen, it has a high degree of non-linearity and does not account for the temperature-dependent *ortho*-*para* ratio of hydrogen. While the Peng–Robinson EOS deviates from experimental data at deep-cryogenic conditions, it provides good agreement over the majority of the cooling range relevant to liquefaction, which brings hydrogen from ambient temperature to near its boiling point. Therefore, the process was modelled using the Peng–Robinson equation of state (EOS) in Aspen Plus. This avoids impractically slow mass- and energy-balance calculations in Aspen and allows for modelling of *ortho*- to *para*-conversion of hydrogen, which is particularly relevant at cryogenic temperatures (Ghafri et al., 2022).

3.4 Conversion Kinetic Model

Following Yin and Ju (2020), the *ortho*-*para* conversion was simplified to be pseudo-first order, i.e. $r_{H_2} = kC_{H_2}$, where r_{H_2} and C_{H_2} are the conversion reaction rate [$\text{kmol m}^{-3}\text{s}^{-1}$] and the concentration of feed hydrogen [kmol m^{-3}], respectively. The reaction is assumed to be catalysed using iron(III) hydroxide, which corresponds to a rate constant k of $1.2 \times 10^3 \text{ s}^{-1}$. As the equilibrium shifts progressively towards *para*-hydrogen with decreasing temperature, the reaction was modelled by a sequence of continuous stirred-tank reactors inserted between the heat exchanger sections, providing a piecewise-continuous approximation of the conversion.

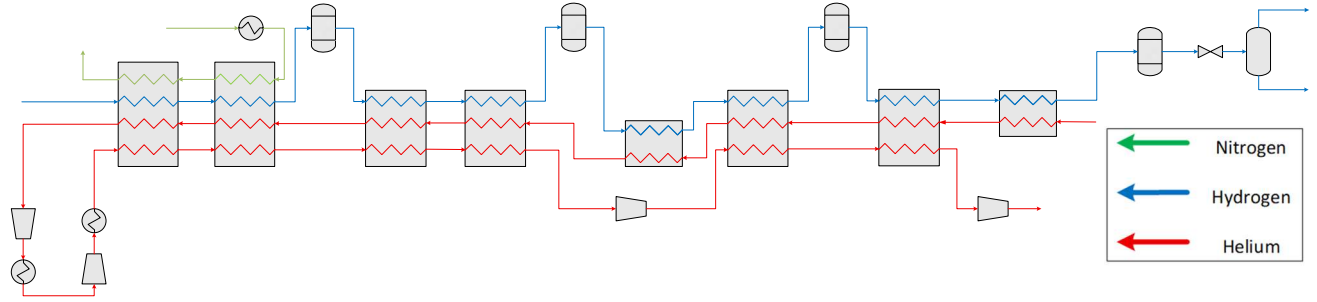


Figure 4: Model of the liquefaction process

3.5 Performance Metric (SEC)

The specific energy consumption (SEC) of a process is commonly used to quantify its energy efficiency. Assessing performance using SEC allows for different designs in this study to be compared consistently. The SEC of the steady-state model can be calculated as follows

$$SEC = \frac{P_{net}}{\dot{m}_{LH_2}} \quad (3)$$

Where P_{net} is the sum of the compressor power requirements and the work produced by expanders, and $\dot{m}_{LH_2}$ is the mass flow rate of liquid hydrogen. Industrial liquefiers typically report SEC values between 10 and 15 kWh/kg_{LH₂} (Ghafri et al., 2022).

3.6 Modelling Approach for Throughput Flexibility

To investigate how the process responds to variations in hydrogen throughput, a flexible process must be designed and rated. As an initial point of consideration, all heat exchanger (HEX) specifications must remain consistent across operating conditions. The process operates with eight heat exchangers: two 2-stream units and six multi-stream heat exchangers (MHEXs), all of which were modelled in Aspen Plus with fixed heat transfer areas to rate their performance and evaluate outlet temperatures.

The 2-stream HEXs were implemented using the standard heat exchanger block. The remaining HEXs were modelled using the MHEX block. To size all HEXs, the steady-state model was simulated with the reference flow rate of 63 kg/h. The calculated areas were then fixed for all exchangers. As Aspen Plus calculates equipment specifications based on steady-state convergence, the heat transfer area is normally recalculated at different flow rates. While the two-stream heat exchanger block allows

for area specification in rating mode, the MHEX block does not allow for an area input.

To fix the heat-transfer area of the MHEXs, design specifications were introduced to keep UA constant (the only area variable accessible on Aspen Plus). Instead of recalculating UA, the design specification fixes it by adjusting the temperature changes of the helium coolant streams (ΔT_C) for all MHEXs. The heat duty (Q) of each HEX is governed by $Q = UA\Delta T_{LM}$. By varying ΔT_C , the temperature driving force ΔT_{LM} is also altered, given its dependence on the coolant's inlet and outlet temperatures. This variation allows the heat exchanger duty Q to change while maintaining a constant UA, allowing throughput flexibility without resizing exchangers. Although liquid nitrogen is present as an additional coolant in the 4-stream MHEXs, helium consistently provides the dominant cooling duty. Therefore, the nitrogen temperature changes are fixed.

To ensure sufficient cooling throughout the HEX network at different hydrogen throughputs, a fixed hydrogen-to-nitrogen flow rate ratio was employed. The nominal hydrogen and nitrogen flow rates in Yin and Ju (2020) are 63 and 660 kg/h, respectively. Therefore, a 1:10.1 ratio was fixed for the rest of the design process. This ratio ensures that the cooling duties in the first two HEXs increase proportionally with hydrogen flow, allowing the temperatures of process streams to remain consistent regardless of hydrogen throughput. On the other hand, the nominal helium flow rate of 1158 kg/h was kept constant, as varying the recycled helium stream introduces significant convergence challenges in Aspen.

Reducing the SEC motivated the consideration of multi-train liquefaction designs. Having multiple smaller trains in parallel would allow each train to be optimised and operated at lower flow rates, with trains turned on in sequence to extend the operating range. Multi-train modelling was first tested by separating the previous

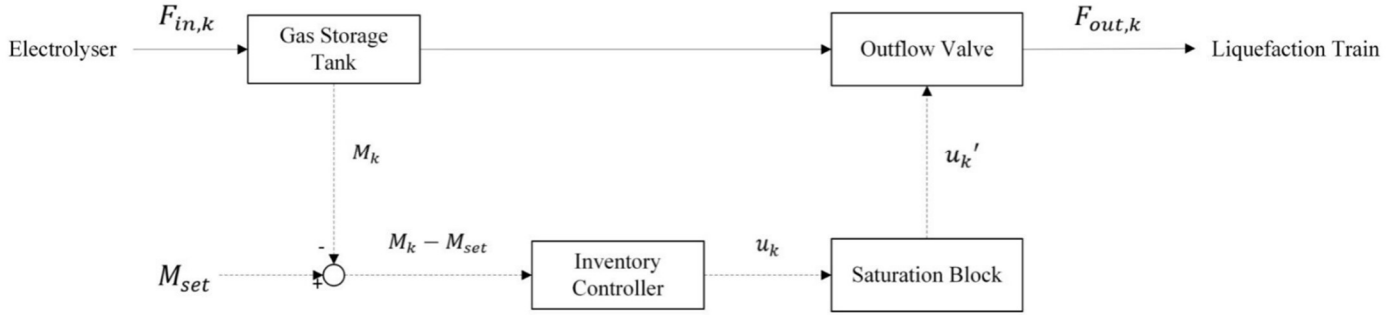


Figure 5: Control block diagram for the gas storage tank

design into two trains. Both trains were optimised for half the original flow rate, 32 kg/h, and all HEXs were resized for the reduced flow rate. The hydrogen-to-nitrogen flow-rate ratio was not changed, while the original helium flow rate was halved to 579 kg/h per train. In addition, a second two-train design was investigated, where the first train was designed for half the original flow rate (32 kg/h), as in the previous design, and the second was designed for a quarter of the original flow rate (16 kg/h). The second train's helium flow rate was reduced to 290 kg/h. For multi-train operation, to achieve the best SEC, the second train is activated when the flow rate exceeds the capacity of the first.

3.7 Reliquefaction Sub-Train Design

After the hydrogen leaves the last HEX, it is expanded through a J-T valve, partially liquefied, and separated into vapour and liquid phases in a flash vessel. Several options were considered for utilising this vapour, including recycling it through the HEX network to recover cooling duty or reconnecting it back into the system before the final cooling stage and expansion. To reduce simulation complexity, the gaseous H_2 byproduct was liquefied in a separate sub-train. As the flash gas is already saturated, it can be liquefied without large power consumption.

The reliquefaction sub-train was modelled on Aspen Plus. The process raises the pressure of H_2 through a cascade of heaters and compressors before expanding it through a J-T valve. The resulting vapour-liquid mixture was then separated in a flash vessel, where the vapour is recycled and fed to the liquefaction plant.

3.8 Hydrogen Gas Storage Control

As wind speed changes over time, the power generated by wind turbines varies accordingly. Consequently, the hydrogen produced by wind-powered electrolysis also fluctuates significantly. Therefore, a gas storage tank was

employed as a buffer, converting the hydrogen supply to a smooth inlet flow for the liquefaction plant. Hourly wind-power data were obtained from Renewables Ninja (2019), and a hydrogen production profile was generated on MATLAB based on the given data. The wind-power data were sourced from a location in Patagonia, selected for its consistently high wind resource. The default turbine capacity, hub height, and turbine model were used, as the dataset is used to generate a normalised variability profile.

3.8.1. Wind-driven hydrogen production profile

Given that the data from Renewables Ninja (2019) are provided at hourly intervals, the sampling time is $\Delta t = 1$ h. Let P_k denote the raw electric power signal at time step k , where $k \in \{1, 2, \dots, N\}$ and N is the total number of time steps. The raw signal is normalised to a dimensionless power output profile p_k ranging from 0 to 1:

$$p_k = \frac{P_k}{\max_{0 \leq j \leq N} P_j} \quad (4)$$

Let $\dot{m}_{H_2, cap}$ denote the effective hydrogen production capacity at $p_k = 1$, i.e. the maximum hydrogen flow rate that can enter the storage tank. All hydrogen flow rates in this section are given in kg/h.

The hydrogen flow rate into the storage tank, $F_{in,k}$, can therefore be defined as follows:

$$F_{in,k} = \dot{m}_{H_2, cap} p_k \quad (5)$$

3.8.2. Tank model and inventory control

A hydrogen gas storage tank was used to decouple the varying hydrogen production from the liquefaction plant inlet. In the model, the tank inventory acts as the controlled variable, and the tank outlet flow is the

manipulated variable used to regulate this inventory. To ensure physically feasible operation, the manipulated outlet flow is limited to the allowable throughput range of the liquefaction plant. For a tank with a chosen capacity M_{max} [kg] and minimum inventory $M_{min} = 0$, its inventory set point, M_{set} , is taken to be half of its capacity, i.e., $M_{set} = 0.5M_{max}$.

The inventory of the gas storage tank, M [kg], is updated using a discrete-time mass balance:

$$M_{k+1} = M_k + (F_{in,k} - F_{out,k})\Delta t \quad (6)$$

Where $F_{out,k}$ is the outflow to the liquefaction plant.

To avoid abrupt throughput changes, the proportional controller in Figure 5 was designed to regulate the tank outflow based on the deviation of the tank inventory from the set point M_{set} :

$$F_{out,k} = F_{nom} + K_c(M_k - M_{set}) \quad (7)$$

Where K_c refers to the proportional controller gain. The controller output is then passed through a saturation block to ensure that the commanded outflow remains within the feasible operating window:

$$F_{min} \leq F_{out,k} \leq F_{max} \quad (8)$$

This represents both the physical limits of the outlet valve and the design throughput limits of the liquefaction plant.

3.8.3. Optimisation of unknown parameters ($\dot{m}_{H_2, cap}$, K_c , M_{max})

The liquefaction model provides the SEC at all integer flow rates in range $[F_{min}, F_{max}]$. Given the controlled outflow profile $F_{out,k}$, the average SEC (SEC_{avg}) of the liquefaction plant is:

$$SEC_{avg} = \frac{\sum_{k=1}^N SEC_k F_{out,k}}{\sum_{k=1}^N F_{out,k}} \quad (9)$$

The decision variables $\dot{m}_{H_2, cap}$, K_c , and M_{max} are then chosen to minimise SEC_{avg} :

$$\min_{\dot{m}_{H_2, cap}, K_c, M_{max}} SEC_{avg} \quad (10)$$

Subject to the following operational constraints:

$$0.1M_{max} \leq M_k \leq 0.9M_{max} \quad (11)$$

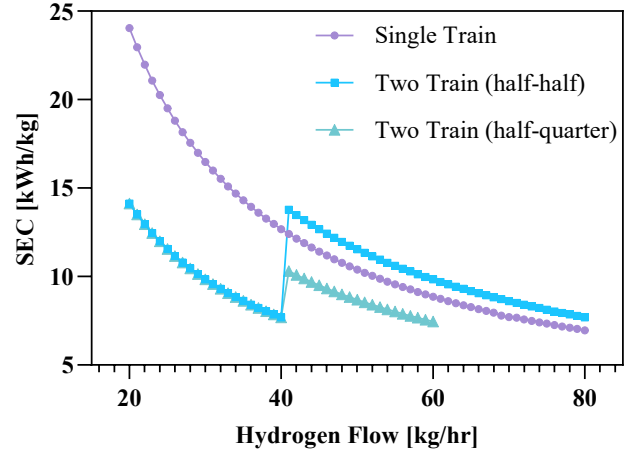


Figure 6: SEC vs Hydrogen Flow for the three designs

$$F_{min} \leq F_{out,k} \leq F_{max} \quad (12)$$

The inventory constraints ensure sufficient inventory for buffering while preventing tank overflow, and the hydrogen outflow constraints ensure the liquefaction plant always operates within its feasible flow rate window. Each decision variable was explored over a wide search range, and a MATLAB-based grid-search optimisation was implemented to identify the most energy-efficient configuration.

4. Results and Discussion

4.1 Varying Throughput Design

A single-train liquefaction process was developed according to the procedure in Section 3.6. The feasible hydrogen flow-rate range was extended from the fixed nominal flow rate of 63 kg/h to a range of 1–80 kg/h. However, this flexibility compromises the energy efficiency of the process, as the SEC values increase sharply when the hydrogen throughput is reduced from the nominal flow rate.

The fixed helium recycle flow rate is the primary contributor to this increase in SEC. The power requirement of the process depends solely on the duties of the helium-cycle compressors. Because the helium flow

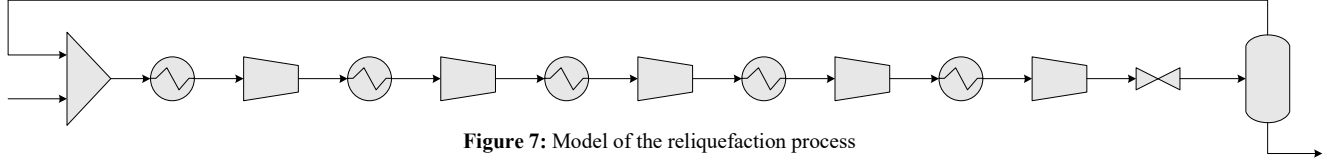


Figure 7: Model of the reliquefaction process

rate is fixed across different hydrogen throughputs and the temperature profiles change very little, P_{net} varies only slightly between operating points. Therefore, because SEC is defined as the energy consumption per mass of liquid hydrogen produced, low hydrogen flow rates result in high SEC values. The first two-train design, which divides the process into two trains of equal capacity, has a feasible hydrogen throughput of 1 to 40 kg/h per train. As a result, the total operating range matches that of the single-train design. When the hydrogen flow rate is between 1 and 40 kg/h, only one of the trains is in operation. Therefore, the total mass flow of helium is halved, reducing the total power consumption and consequently the SEC, as shown in Figure 6. However, when the hydrogen flow rate exceeds 40 kg/h, both trains must be in operation to support feasible liquefaction, so the total mass flow of helium is the same as that of the single-train configuration. Therefore, the energy performance of this two-train design declines sharply after the second train is turned on.

The second two-train design addresses this issue by halving the operating range of one of the liquefaction trains, i.e., one train operates for hydrogen flow rates of 1 to 40 kg/h, while the other has an operating range of 1 to 20 kg/h. Given that the second train operates at one-quarter of the original helium flow rate, the total mass flow of helium for this configuration never exceeds 75% of that of the single-train model. Therefore, this design produces the lowest SEC values across its operating range, as illustrated in Figure 6.

When comparing the SEC values of all three designs, both two-train designs exhibit an increase in SEC when the second train turns on, as the liquefaction trains are switched on in sequence. This abrupt increase in SEC creates a sharp transition, or ‘ramp-up,’ when the second train activates, which can be challenging to control in practice. However, the second two-train design, which has a feasible throughput range of 1-60 kg/h, has a smoother SEC jump than the other two-train configuration due to its lower power consumption. Therefore, the second two-train design was considered the optimal configuration, given its lower SEC and smoother ramp-up response.

4.2 Reliquefaction Sub-Train

The final design of the reliquefaction train is shown in Figure 7. The flash gas leaves the main train near the saturation point (1.5 bar, 21.75 K). The sub-train brings the stream pressure up to 14.1 bar before expanding it down to 1.5 bar. A minimum of 5 heaters and compressors were required to bring the stream up to a pressure where its expansion would lead to partial liquefaction. The final J-T expansion allows the fluid to reach a vapour fraction of 0.97. As previously mentioned, the power consumption in the main process does not vary greatly with changes in hydrogen flow rate. In the reliquefaction train, as hydrogen is compressed, the power consumption depends solely on its throughput. As the flow rate increases, the power requirement of reliquefaction increases. Although power consumption in the sub-train increases with higher flow rates, the sub-train recovers more liquid hydrogen in the process. For example, when operating at a hydrogen throughput of 63 kg/h, the single-train model only recovers 17.38 kg/h of liquid hydrogen, while the vapour product is released as flash gas. Without the liquefaction, the SEC is calculated as 25.80 kWh/kgLH₂, whereas reliquefaction reduces the SEC down to 8.50 kWh/kgLH₂. Therefore, the increase in liquid hydrogen flow greatly compensates for the increase in total power requirement, highlighting the necessity of reliquefaction.

4.3 Hydrogen Gas Storage and Overall Energy Performance

The optimisation identified an effective hydrogen production capacity of 70 kg/h, with a storage requirement of 850 kg and a proportional gain of 0.12 h⁻¹. The tank size is operationally realistic, as modern compressed-hydrogen tube trailers typically store 500 to 900 kg of hydrogen (Brown, 2019). Under this configuration, the liquefaction plant operated at a mean inlet flow of 17.65 kg/h and achieved an average SEC of 16.47 kWh/kgLH₂. These

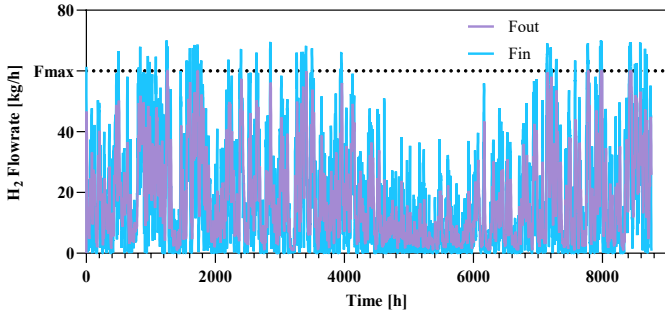


Figure 8: Change in Hydrogen Flow rate over time

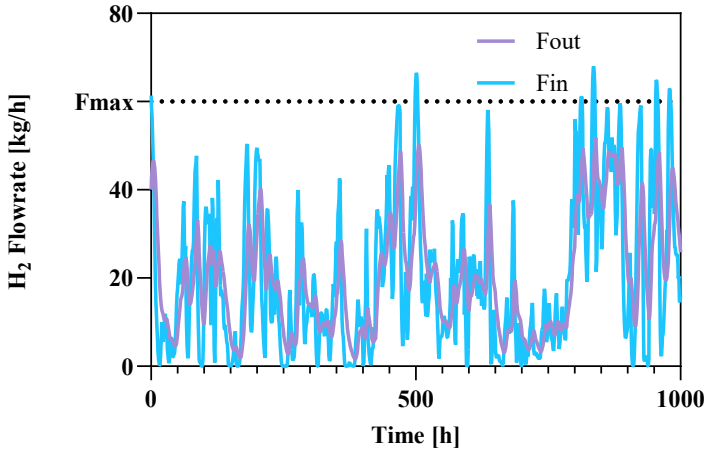


Figure 9: Change in Hydrogen Flow rate over time (first 1000 hours)

values represent the best compromise within the chosen search space between hydrogen production sizing, storage capacity, aggressiveness, and the specific energy consumption of the downstream liquefaction plant.

4.3.1 Hydrogen inflow vs. controlled outflow

The inflow profile ($F_{in,k}$), derived from normalised wind-power data, shows rapid fluctuations between 0 and 70 kg/h. The controlled outflow ($F_{out,k}$), however, remains within the feasible operating range of 1 - 60 kg/h and exhibits reduced variability in Figure 9. This demonstrates that the controller effectively dampens large oscillations, allowing the liquefaction plant to operate without infeasible spikes in throughput.

4.3.2 Tank Inventory Behaviour

The inventory mass varies between 94 kg and 748 kg, well within the imposed bounds of 10–90% of capacity (85–765 kg), which indicates that the tank never failed to supply hydrogen during low-wind periods, while surplus hydrogen during high-wind periods was successfully stored in the tank. This confirms that the chosen capacity of 850 kg is indeed adequate for the given wind profile.

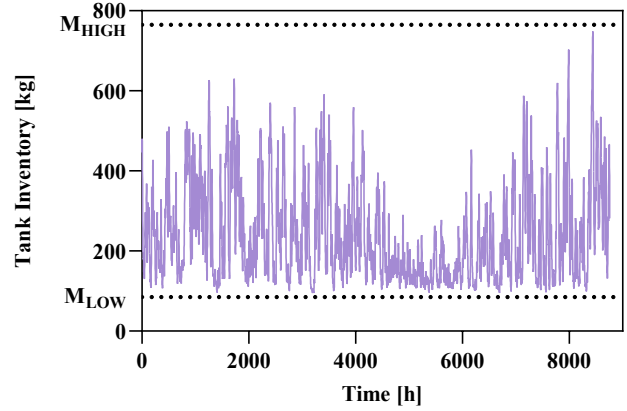


Figure 10: Change in Tank Inventory over time

4.3.3 Impact on Liquefaction Energy Consumption

As shown in Figure 6, the specific energy consumption (SEC) of the two-train liquefaction plant is highest when either train operates at its minimum operable load. This occurs at $F_{out,k} = 1$ kg/h, i.e. one train at its minimum load, or $F_{out,k} = 41$ kg/h, i.e. the first train operates at its maximum load at 40 kg/h, and the second train operates at 1 kg/h. Therefore, the minimisation of SEC implicitly favours a smooth $F_{out,k}$ profile, as a highly oscillatory $F_{out,k}$ profile would reach the minimum load frequently, thereby heavily increasing the average SEC of the liquefaction plant.

4.4 Limitations

4.4.1 Ramp-Up

To minimise the SEC of the optimal two-train configuration in Section 4.1, the second train—designed for a smaller throughput range of 1–20 kg/h—only switches on when the hydrogen flow rate exceeds 40 kg/h, i.e., the maximum operable flow rate of the first train, as shown in Figure 6. Consequently, when the hydrogen flow rate fluctuates around 40 kg/h, this configuration becomes operationally infeasible, as it would lead to frequent start-up and shutdown of the second liquefaction train, imposing substantial mechanical stress on the process equipment.

4.4.2 Liquid Nitrogen Power Requirement

The liquid nitrogen supplied to the liquefaction plant undergoes phase change and leaves the process as nitrogen gas, so it is not recycled. Therefore, liquid nitrogen must be supplied continuously. Industrial hydrogen liquefiers are often built in proximity to air separation plants, where

liquid nitrogen is readily available. The SEC evaluated in this study accounts only for the power requirements of the hydrogen liquefaction plant and the reliquefaction sub-train, which implicitly assumes that the facility is located near an air separation plant. However, for liquefaction plants where liquid nitrogen is not readily available, the power required to produce liquid nitrogen should be included in the total power consumption of the process, which would increase the SEC. Therefore, the SEC reported in this study is likely underestimated.

4.4.3 Tank Modelling Limitations

Although the optimised configuration achieves a smooth hydrogen throughput, the resulting average SEC (16.47 kWh/kg_{LH2}) remains higher than that of state-of-the-art liquefaction technology, which typically requires 12 kWh/kg_{LH2} (Essler et al., 2012). This is because the wind resource exhibits periods of low output, which limits the plant to lower hydrogen flow rates. While increasing the pre-feed storage tank size could reduce downtime, current limitations on the gaseous hydrogen capacity of trailer tanks make this operationally infeasible. As a result, the liquefaction plant must sometimes operate below the optimal throughput of 40 kg/h, increasing the average SEC despite the smoothing effects of the controller.

Beyond these system-level constraints, the simplifications to the model also limit the fidelity of the results. The model treats the tank as a mass accumulator, which does not account for variations in pressure and temperature during operation. As a result, the SEC of the model is likely underestimated, because additional expanders, compressors, and heat exchangers are likely required to control pressure and temperature in the tank. The model also assumes perfect instantaneous responses to wind fluctuations, which overestimates the achievable smoothness and underestimates how often the tank might reach operational limits. The optimisation is performed using a discrete grid search and is limited to one year of wind data, without incorporating economic cost trade-offs or multi-year variability. While not intended for operational control, the model serves as an early-stage conceptual design, providing valuable insights into trade-offs between storage capacity, production sizing, and SEC performance.

5. Conclusion

This study demonstrates that fluctuations in wind-powered hydrogen production can be effectively mitigated using a gas storage buffer. A simple feedback control strategy, which adjusts hydrogen outflow based on tank inventory, enables the storage tank to smooth variations in hydrogen throughput and maintain stable operation of both the tank and the downstream liquefaction plant. To model a flexible liquefaction process, multiple designs were tested on Aspen Plus, in which helium-side temperature differences were adjusted while heat-exchanger areas were fixed. The liquefaction train development resulted in a final design that operates two trains in parallel. The gaseous hydrogen byproduct from the main train was liquefied separately in a sub-train. This approach allows realistic operation over hydrogen flow rates between 1 and 60 kg/h. The final optimised storage and liquefaction design operates at an average SEC of 16.47 kWh/kg_{LH2}, which is comparable with current hydrogen liquefaction plants.

6. Acknowledgements

We thank our GTA, Cameron Aldren, for his continued advice and insight throughout this project.

References

- Alanazi, K., Mittal, S., Hawkes, A. and Shah, N. (2025). Renewable hydrogen trade in a global decarbonised energy system. *International Journal of Hydrogen Energy*, [online] 101, pp.712–730. doi:<https://doi.org/10.1016/j.ijhydene.2024.12.452>.
- Brown, A. (2019). Hydrogen Transport. [online] www.thechemicalengineer.com. Available at: <https://www.thechemicalengineer.com/features/hydrogen-transport/>.
- Essler, J., Haberstroh, C., Quack, H., Walnum, H.T., Berstad, D., Neks , P., Stang, J., Decker, L., Treite, P., B rsch, M. and Holdener, F. (2012). Integrated design for demonstration of efficient liquefaction of hydrogen (IDEALHY) [online] Available at: https://www.idealhy.eu/uploads/documents/IDEALHY_D1-1_Report_Tech_Overview_and_Barriers_web2.pdf.
- Ghafri, S.Z.A., Munro, S., Cardella, U., Funke, T., Notardonato, W., Trusler, J.P.M., Leachman, J., Span,

R., Kamiya, S., Pearce, G., Swanger, A., Rodriguez, E.D., Bajada, P., Jiao, F., Peng, K., Siahvashi, A., Johns, M.L. and May, E.F. (2022). Hydrogen liquefaction: a review of the fundamental physics, engineering practice and future opportunities. *Energy & Environmental Science*, [online] 15(7), pp.2690–2731. doi:<https://doi.org/10.1039/D2EE00099G>.

Hoecke, L.V., Laffineur, L., Campe, R., Perreault, P., W. Verbruggen, S. and Lenaerts, S. (2021). Challenges in the use of hydrogen for maritime applications. *Energy & Environmental Science*, [online] 14(2), pp.815–843. doi:<https://doi.org/10.1039/D0EE01545H>.

Koons, E. (2023). The Pros and Cons of Hydrogen Energy. [online] *Energy Tracker Asia*. Available at: <https://energytracker.asia/pros-and-cons-of-hydrogen-energy/>.

Linde-engineering.com. (2021). Hydrogen in Leuna: The Success Story Continues. [online] Available at: <https://www.linde-engineering.com/products-and-services/success-stories/2024/hydrogen-in-leuna-the-success-story-continues>.

Liu, Z., Deng, Z., Davis, S. and Ciais, P. (2023). Monitoring global carbon emissions in 2022. *Nature Reviews Earth & Environment*, 4. doi:<https://doi.org/10.1038/s43017-023-00406-z>.

A. Maczek and A. Meijer. *Statistical Thermodynamics*. 2nd. Oxford, England: Oxford University Press, 2017.

National Institute of Standards and Technology (2023). Isobaric Properties for Hydrogen. *NIST Chemistry Webbook*, [online] 69. doi:<https://doi.org/10.18434/T4D303>.

Renewables Ninja (2019). *Renewables.ninja*. [online] *Renewables.ninja*. Available at: <https://www.renewables.ninja/>.

Yin, L. and Ju, Y. (2020). Process optimisation and analysis of a novel hydrogen liquefaction cycle. *International Journal of Refrigeration*, 110, pp.219–230. doi:<https://doi.org/10.1016/j.ijrefrig.2019.11.004>.

Investigation of different factors affecting CO₂ capture by CALF-20

Zhixiang Huang & Haotian Rong

Department of Chemical Engineering, Imperial College London, UK

Abstract

This work investigates how shaping CALF-20 into pellets and adding bentonite as a binder influence its structural stability and CO₂ adsorption behavior. Once the CALF-20 powder was obtained, it was shaped into pellets under two different pressing loads, 0.4 T and 0.6 T. Two more pellets containing 10 wt% and 20 wt% of bentonite were also made, so that the influence of the binder could be compared directly. Commercial zeolite 13X pellets were included simply as a reference point. Structural properties were examined by XRD and FTIR, while SEM was used to observe surface morphology and particle structure. From the BET results, pelletization itself did not cause any obvious loss of microporosity in CALF-20. However, once bentonite was incorporated, both the surface area and the micropore volume declined, largely because the binder has very limited porosity and effectively dilutes the active framework. CO₂ uptake was measured by TGA at 303 K. As expected, the samples containing more bentonite showed a drop in adsorption capacity, which matches the reduction in micropore volume. These results help clarify the way CALF-20 responds to different pressing pressures and binder amounts. Taken together, the findings show quite clearly how the pressing conditions and the binder content influence CALF-20. This information should be helpful for improving or tailoring pellet-type adsorbents for CO₂ capture.

Keywords: CALF-20, zeolite 13X, bentonite, pellets, CO₂ uptake, adsorption

1 Introduction:

In recent years, rising carbon dioxide emissions have been recognized as a major factor contributing to global warming [1]. This has led to growing interest in identifying practical and efficient ways to reduce CO₂ output. Among the available mitigation approaches, solid adsorption has gained increasing interest, largely because of the low energy of regeneration and high selectivity for CO₂, it avoids the corrosive issues associated with some liquid absorbents, and it is generally viewed as a more environmentally benign option. The fundamental principle of solid adsorption is that CO₂ molecules from a gas stream are captured through physisorption (through van der Waals interactions), chemisorption (gas molecules form chemical bonds with specific active sites on the adsorbent), or a combination of both [2]. Despite these benefits, many commonly used solid adsorbents still have notable drawbacks, especially their poor stability when exposed to moisture and the high raw material cost. These issues make it clear that we still

need materials capable of achieving high CO₂ uptake while remaining structurally stable under conditions closer to real operation. One material that has recently drawn attention in this regard is CALF-20, a zirconium-based MOF that shows considerable potential for overcoming these challenges [3]. The metal organic framework (MOFs) material like CALF-20 has high selectivity and shaping plays an important role in practical cases. Additionally, binders are often used to increase the structural stability, but it will cause the block of pores, reduce surface area which result in reduction in adsorption capacity [4]. Therefore the research should not only focus on the material itself but also on optimal shaping conditions for more practical uses.

In this paper, we begin by synthesizing CALF-20 and subsequently investigate how different binder ratios affect the properties of samples, including CO₂ adsorption capacity, structural stability, pore-size distribution, and other relevant characteristics.

2 Experimental Framework

2.1 Materials:

Zeolite 13X powder (CAS# 63231-69-6) and bentonite powder (CAS# 1302-78-9) were provided by Sigma-Aldrich, and methanol (CAS# 67-56-1) were provided by VWR, 1,2,4 triazole powder (CAS# 288-88-0) and zinc oxalate dihydrate (CAS# 55906-21-3) were provided by Thermo Fisher Scientific. We use N₂ (zero grade, 99.9995%), He (CP grade, 00.9990%) and CO₂ (research grade, 99.999%) gases in this study. All chemicals and materials were used in their original form unless stated otherwise.

The crystalline phase was determined by using Malvern Panalytical Aeris X-ray Diffractometer. XRD patterns were collected over a scan range of 5°–90° (2θ) with a step size of 0.01° and a counting time of 29 s per step. XRD pattern was conducted with powder

The particle size and morphology of CALF-20 were observed by Zeiss Sigma 300 SEM.

The functional group and chemical bonding environment were determined by FTIR002E

2.2 Synthesis CALF-20 powder:

6.60 g of zinc oxalate, 5.00 g of 1,2,4-triazole and 66.0 mL of methanol were added into a 125 mL Teflon autoclave. Leave the autoclave in the oven at 180°C for 48 hours. The product was washed with methanol [5]. After that, the product will be subjected to centrifugation to achieve solid–liquid separation and then placed in a fume cupboard for 48 hours to allow the methanol to evaporate. If the powder remains wet, it can be further dried in a drying oven.

2.3 Preparation of bentonite, zeolite 13X and CALF-20 pellets:

To investigate the differences between pellets formed from pure CALF-20 powder under different compression pressures, two pellets were prepared at 0.4 ton and 0.6 ton respectively. Additionally, to investigate the effect of

different binder ratios on the samples, we used bentonite at binder ratios of 10% and 20%, and compressed them with CALF-20 into pellets under a pressure of 0.4 ton. We also prepared zeolite 13X pellets under a compression pressure of 0.4 ton to compare their CO₂ adsorption capacity with that of CALF-20. The sample table is listed in the appendix A.1.

2.4 Materials characterization

2.4.1 BET

BET surface area, total pore volume, and micropore volume are measured by Tristar 3000 to evaluate the porous structure of the materials and mesopore volume is calculated from total pore volume and micropore volume, as these parameters directly influence the available adsorption sites, gas diffusion pathways, and ultimately the CO₂ adsorption capacity and kinetics.

Mesopore volume can be estimated by the following equation [6]:

$$V_{meso} = V_{tot} - V_{micro} \quad (1)$$

Where V_{meso} , V_{tot} and V_{micro} are mesopore volume, total pore volume and micropore volume respectively. Prior to the measurements, the samples were degassed ex-situ at 120 °C for at least 24 hours to ensure complete removal of adsorbed gases.

2.5 TGA

CO₂ adsorption experiment was carried out on a TGA 550 at 303 K and 1 atm. Prior to the CO₂ measurements, the samples were degassed in-situ at 393 K. They were heated from room temperature to 120 °C at a rate of 10 °C·min⁻¹ and then held isothermally for one hour to ensure complete degassing. After that, the sample was cooled inside the instrument to 30 °C at a rate of 10 °C·min⁻¹ and held isothermally for 10 minutes. This step is necessary because high or fluctuating temperatures can lead to inaccurate CO₂ uptake measurements. 30 °C is selected to align with the industry standard for solid sorbent capture

since this temperature maximizes the adsorption capacity [7]. Finally, a gas mixture consisting of 15% CO₂ and 85% N₂ was passed over the sample at a total flow rate of 100 mL·min⁻¹, and the sample was maintained at 30 °C for one hour to measure the change in mass.

The CO₂ uptake can be calculated using the following equation:

$$CO_2 \text{ Uptake} \left(\frac{\text{mmol}}{\text{g}} \right) = \frac{W_{\text{Final}} - W_{\text{Before } CO_2 \text{ flowing}}}{W_{\text{sample}}} \quad (2)$$

Where: W_{Final} is the weight of final reading, $W_{\text{Before } CO_2 \text{ flowing}}$ is the weight before the gas mixture is flowing and W_{sample} is the weight of after heat to 120 °C and held isothermally for one hour.

3 Results and Discussion

3.1 CALF-20 characterisation by FTIR and XRD

To verify the chemical structure and crystalline structural integrity of the synthesized adsorbents, Fourier-transform infrared (FTIR) spectroscopy and X-ray diffraction (XRD) were conducted on CALF-20 and zeolite 13X powders and the results were shown on Figure 3.1.1 and Figure 3.1.2 below.

The FTIR spectrum on Figure 3.1.1 of CALF-20 powder exhibits characteristic vibrational peaks that are mostly consistent with the pristine CALF-20 framework, including ligand-derived C-N stretching modes in the range around 1450-1550 cm⁻¹ and C-O bending modes 820-880 cm⁻¹, confirming its intrinsic functional groups were shown compared to the literature data [8]. Two extraneous signals were also observed in the CALF-20 spectrum: a peak centered at around 1100 cm⁻¹, which is assigned to the C-O stretching vibration of residual methanol and indicates incomplete solvent removal during post-synthesis drying, and an unassigned peak at around 650 cm⁻¹ that does not correspond to CALF-20 characteristic vibrational modes or the residual solvents

evaluated in this work. For zeolite 13X powder, the FTIR spectrum on Figure 3.1.1 showed the aluminosilicate framework vibrations, including T-O-T (T = Si/Al) stretching modes in the range of around 1000 cm⁻¹ and bending modes 450-550 cm⁻¹ which matches the data of pristine zeolite 13X from literature [9]. No extraneous peaks were detected in the zeolite 13X spectrum which verified the structural integrity of its aluminosilicate framework.

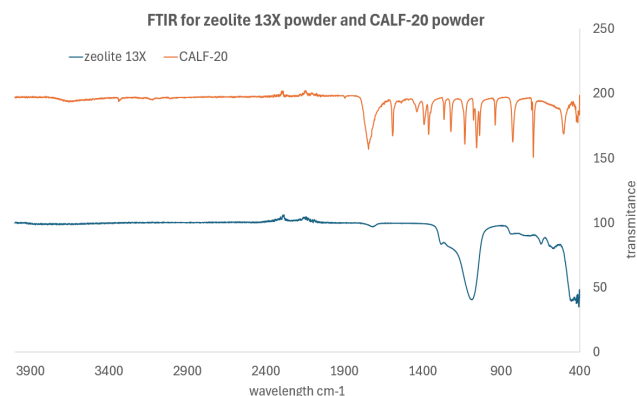


Figure 3.1.1: FTIR for zeolite 13X powder and CALF-20 powder

The XRD pattern on Figure 3.1.2 of CALF-20 powder displayed a sharp, intense characteristic diffraction peak at around 15°, which is consistent with literature-reported data for CALF-20 XRD spectrum and some weaker secondary peaks that match the standard XRD profile of CALF-20 [10]. The absence of extraneous crystalline diffraction peaks confirms the high purity of the synthesized CALF-20. For zeolite 13X powder the XRD pattern on Figure 3.1.2 exhibited has characteristic diffraction peaks, for instance at around 25°, which matches with the literature data for zeolite 13X [11]. Although the XRD pattern of our zeolite 13X shows some differences compared with the literature, the FTIR spectrum matches very well with the reported data. This indicates that the types of chemical bonds and functional groups in our material are consistent with those of the zeolite 13X examined in the reference. The different in the XRD patterns reflect differences in crystal structure, and the most common reason for such variations is differences in crystallinity. Even products manufactured by the same supplier can exhibit variations in crystallinity, and different synthesis route, operating conditions and raw

material used can further lead to different XRD results. At the same time, several diffraction peaks in our sample still coincide with those in the literature—for example, the peak slightly above 5° , the two peaks around 10° , the peak near 15° , the peak at approximately 22.5° , the peak slightly above 25° , and the peak just below 30° . These consistent features further confirm that the majority of the crystalline framework is the same as the reported zeolite 13X. The remaining differences may arise from variations in synthesis route, processing conditions, and raw material between our zeolite 13X supplier and literature. All of them directly influence crystallinity, crystal size, preferred orientation, and the Si/Al ratio. These factors have a directly influence on XRD patterns. Overall, these FTIR and XRD results confirm the successful synthesis of CALF-20, with their chemical structures. Meanwhile, we can be reasonably confident that the prepared sample is zeolite 13X but with different crystal structure with the zeolite 13X used in the literature.

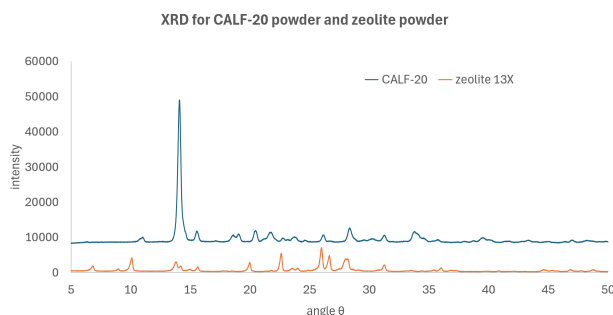


Figure 3.1.2 XRD spectrum for CALF-20 powder and zeolite 13X powder

3.2 BET analysis

Nitrogen adsorption–desorption measurements were performed to evaluate the textural properties of the prepared samples. In this study, the total pore volume (V_{tot}) was obtained by the value of quantity of liquid nitrogen adsorbed from the adsorption isotherm profile at $p/p_0 = 0.97$, while the micropore volume (V_{micro}) was determined by directly reading the micropore volume from the t-plot report and the data is shown in table 3.2.1. The mesopore volume (V_{meso}) was calculated as the difference between V_{tot} and V_{micro} . As expected, bentonite powder and bentonite pellets exhibited very low

BET surface areas ($21\text{--}22 \text{ m}^2\cdot\text{g}^{-1}$), with both micropore and mesopore volumes being negligible, confirming that bentonite behaves as a non-porous diluent in the composite pellets. In contrast, CALF-20 powder showed a significantly higher BET surface area of $365 \text{ m}^2\cdot\text{g}^{-1}$ together with a micropore volume of $0.168 \text{ cm}^3\cdot\text{g}^{-1}$ and a measurable mesopore contribution. For comparison, commercial zeolite 13X pellets also exhibited a high BET surface area (approximately $490 \text{ m}^2\cdot\text{g}^{-1}$) and substantial microporosity, consistent with its characteristic crystalline aluminosilicate framework. Pelletization of CALF-20 at 0.4 T and 0.6 T resulted in surface areas of $350 \text{ m}^2\cdot\text{g}^{-1}$ and $334 \text{ m}^2\cdot\text{g}^{-1}$, respectively, with only slight decreases in both micropore and mesopore volumes. These results indicate that pelletization causes minimal structural collapse of the CALF-20 pore network and preserves most of its accessible porosity across both micropore and mesopore ranges. When bentonite was introduced as a binder, the BET surface area decreased to $312 \text{ m}^2\cdot\text{g}^{-1}$ (10 wt%) and $303 \text{ m}^2\cdot\text{g}^{-1}$ (20 wt%), accompanied by corresponding reductions in both micropore and mesopore volumes. This reduction is primarily attributed to the dilution effect of the low-porosity bentonite phase rather than pore blockage, as the microporous fraction remains dominant and scales proportionally with the CALF-20 content. Furthermore, the mesopore volumes remain small relative to the total pore volume for all samples, confirming that CALF-20 is predominantly microporous and that neither pelletization nor binder addition generates significant mesoporosity. Overall, the BET results demonstrate that CALF-20 retains its porous structure after pelletization, and although the incorporation of bentonite reduces the apparent surface area and pore volume per gram of pellet, the intrinsic microporosity of CALF-20 is largely preserved.

samples	BET surface m^2/g	$V_{tot} cm^3/g$	$V_{meso} cm^3/g$	$V_{micro} cm^3/g$
bentonite pellet 0.4 T	22	32.445	32.442	0.003
bentonite powder	21	38.175	38.173	0.002
CALF-20 with 10%wt bentonite 0.4 T	312	100.609	100.462	0.147
CALF-20 with 20%wt bentonite 0.4 T	303	98.130	97.992	0.138
CALF-20 powder	365	110.559	110.391	0.168
CALF-20 pellet 0.4 T	350	109.896	109.737	0.159
CALF-20 pellet 0.6 T	334	109.919	109.769	0.150
zeolite 13X pellet	420	133.362	133.189	0.173
zeolite 13X powder	462	146.121	145.933	0.188

Table 3.2.1 BET surface area, total volume of pores, mesopore volume and micropore volume that is determined by physisorption at 77K

3.3 SEM

The SEM micrograph of the synthesized CALF-20 on figure 3.3.1 shows micron-sized crystalline fragments with irregular, plate-like morphology. The particles exhibit rough surfaces and form loosely packed aggregates. No abnormal structures or impurities were observed.

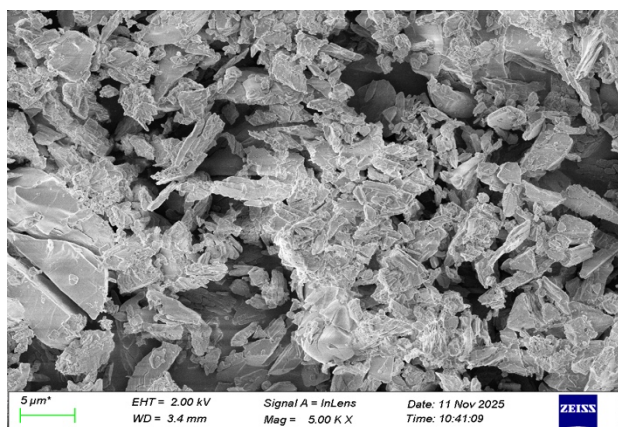


Figure 3.3.1 SEM image of CALF-20 powder

3.4 CO₂ uptake test

Thermogravimetric analysis (TGA) was conducted for all

synthesized samples to quantitatively measure their CO₂ adsorption capacities which were focused on isolating and quantifying physisorptive contributions, the residual volatile impurities are controlled during measurement [12]. Figure 3.4.1 depicts the dynamic profile of typical sample mass change and corresponding temperature ramp for a CALF-20/bentonite composite pellet (20 wt% bentonite), both the pre-adsorption purification step and subsequent CO₂ uptake step is recorded. During the initial temperature elevation segment, a gradual mass loss was observed that is due to the desorption of water and residual methanol solvent entrained within the composite pores during synthesis. The subsequent isothermal period yielded a stable mass plateau which confirmed the complete elimination of these volatile contaminants at which a critical pretreatment step to prevent interference with subsequent CO₂ adsorption measurements. Following purification, the system was cooled to the target adsorption temperature while starting uninterrupted CO₂ flow, prompting a monotonic mass increase in the composite, this mass gain persisted until a steady-state plateau was achieved, this indicates that the material has reached adsorption equilibrium and the accessible active sites have been saturated with CO₂ [13].

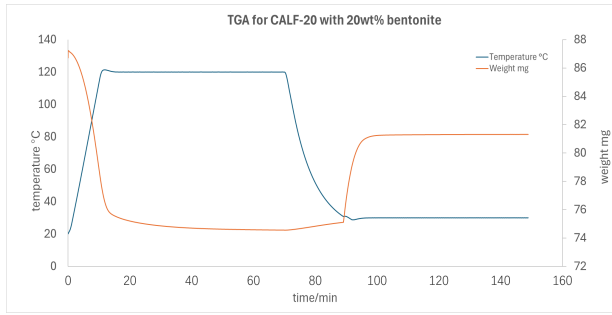


Figure 3.4.1 TGA profile for CALF-20 with 20wt% bentonite

The Figure 3.4.2 illustrates the CO₂ adsorption capacities of all the prepared adsorbent samples under the identical experimental conditions. Notably, for the same base adsorbent material, powder-form samples exhibit much higher CO₂ uptake than their pelletized samples. For instance, zeolite 13X powder achieves a CO₂ adsorption capacity of 1.79 mmol·g⁻¹, which is higher than that of pelletized zeolite 13X. Furthermore, CALF-20 demonstrates the most superior CO₂ adsorption performance among all tested materials. For example, CALF-20 powder attains a CO₂ uptake of approximately 2.03 mmol·g⁻¹, outperforming the 1.79 mmol·g⁻¹ capacity of zeolite 13X powder by roughly 13.4%.

To indicate the effect of bentonite loading on CO₂ adsorption behavior, pure bentonite and CALF-20/bentonite mixed powder composites were compared for adsorption tests. Experimental results reveal a monotonic reduction in CO₂ adsorption capacity with the increase of bentonite content in CALF-20-based composites. This adverse trend can be attributed to the bentonite-induced blockage of active micropores and mesopores within CALF-20, which restricts the diffusion of CO₂ molecules to intrinsic adsorption sites and thereby impairs the overall uptake capability of the composite. These observations are in good agreement with the BET nitrogen adsorption characterization results which shows a positive correlation between the total pore volume of the adsorbent and its CO₂ adsorption capacity. In simple term, adsorbents with larger total pore volumes exhibit better

CO₂ uptake efficiency.

Finally, the impact of compaction pressure on the adsorption performance of pelletized CALF-20 was investigated by making pellets under two distinct pressures (0.4 ton and 0.6 ton). The CO₂ adsorption capacities of these two pellet samples do not show statistically significant differences which indicates that compaction pressures in the range of 0.4–0.6 ton do not cause significant changes to the pore structure and accessible adsorption sites of CALF-20, so the absorption behavior of CO₂ is similar.

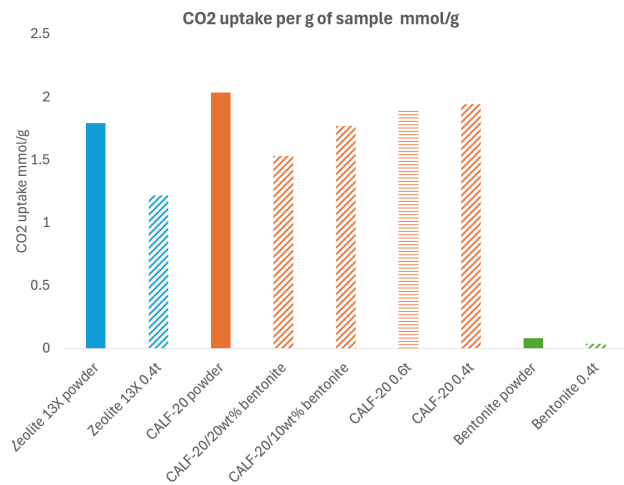


Figure 3.4.2 Amount of CO₂ uptake for all prepared samples

To establish a quantitative correlation and estimation between the CALF-20/bentonite mass ratio and CO₂ adsorption performance, Figure 3.4.3 presents a direct comparison between experimentally measured CO₂ uptake values and theoretically calculated CO₂ uptake values for the prepared composites. For the CALF-20 composites loaded with 10 wt% and 20 wt% bentonite, the experimentally determined CO₂ uptake capacities are in excellent agreement with the theoretical predictions with negligible deviations observed between the two data sets. The calculated data is by equation 3:

$$U \left(\frac{\text{mmol}}{\text{g}} \right) = x_{\text{CALF-20}} \times U_{\text{CALF-20}} + x_{\text{bentonite}} \times U_{\text{bentonite}} \quad (3)$$

Where U is amount of CO₂ uptake, $x_{\text{CALF-20}}$ and

$x_{\text{bentonite}}$ are separately the percentage of material in pellet. The CO₂ uptake for CALF-20 pellet and bentonite pellet is $U_{\text{CALF-20}}$ and $U_{\text{bentonite}}$.

This strong consistency validates the reliability of the calculation model which enables it to predict the carbon dioxide adsorption performance of CALF-20/ bentonite composites with a different bentonite mass ratio.

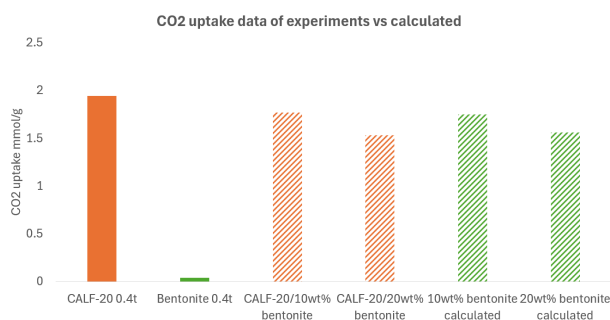


Figure 3.4.3 CO₂ uptake amount from experiments vs calculated

4 Conclusion

In this work, CALF-20 were successfully synthesized and characterized, and CO₂ adsorption performance was systematically evaluated for both CALF-20 and zeolite 13X, with bentonite used as a binder to make composite pellets for commercial purpose.

First, structural characterization via FTIR and XRD confirmed the integrity of the target adsorbent materials. The FTIR spectrum of CALF-20 showed C-N and C-O peak, with only small amount of residual methanol signals and one undefined peak that did not affect the main structure; zeolite 13X spectrum showed typical aluminosilicate T-O-T peak, verifying its structural stability. XRD patterns further validated the synthesis of CALF-20 and zeolite 13X, without other crystalline impurities. SEM observations also revealed that CALF-20 possessed the typical microcrystalline morphology reported in the literature, confirming successful synthesis.

Second, BET textural property analysis indicated the impact of pelletization and bentonite addition on CALF-20 porous structure. Bentonite itself has a negligible porosity from its low BET surface and micropore volume value which acts as a pore-free diluent rather than a pore-blocking agent. CALF-20 powder showed a high BET surface area of 365 m²·g⁻¹ and micropore volume of 0.168 cm³·g⁻¹, and pelletization of CALF-20 under 0.4–0.6 ton only caused slight reductions in its porous properties which indicates framework collapse is small. The introduction of 10–20 wt% bentonite resulted in decrease in surface area of CALF-20 and micropore volume which revealed the dilution effect of bentonite rather than pore blockage, as the microporous fraction remained dominant with CALF-20 content.

Third, TGA-based CO₂ uptake tests showed that CALF-20 was better than zeolite 13X in CO₂ adsorption. Powder-form adsorbents consistently exhibited higher CO₂ uptake performance than their pelletized form for the same material, due to reduced mass transfer resistance in powder samples. Bentonite loading led to a constant decrease in CO₂ adsorption capacity which is correlated with the total pore volume of the composite which is consistent with BET results. Additionally, compression pressures of 0.4–0.6 ton had little significant impact on the CO₂ uptake of CALF-20 pellets, as the pore structure and adsorption sites were kept under these pressures. The established linear mixing model also showed excellent consistency between calculated and experimental CO₂ uptake data for 10 and 20 wt% bentonite composites which enable reliable estimation of the performance of composites with higher bentonite ratios.

All in all, this work confirms that CALF-20 is a high-performance CO₂ adsorbent, also shows bentonite can serve as a feasible binder to make pellets with little loss in adsorption capacity for CO₂ capture. From our

experiments, we can conclude that pellets prepared with a 20% binder ratio of bentonite are already mechanically strong, whereas those made with a 10% binder ratio still show insufficient strength. Furthermore, future research could focus on optimizing the bentonite-CALF-20 ratio, different binders and pelletization conditions to balance mechanical strength and adsorption performance to make it more suitable for commercial purposes.

References :

[1]

L. J. R. Nunes, "The rising threat of atmospheric CO₂: A review on the causes, impacts, and mitigation strategies," MDPI, <https://www.mdpi.com/2248060> (accessed Dec. 10, 2025).

[2]

F. Raganati, F. Miccio, and P. Ammendola, "Adsorption of carbon dioxide for post-combustion capture: A review," *Energy & Fuels*, vol. 35, no. 16, pp. 12845–12868, Aug. 2021, doi: 10.1021/acs.energyfuels.1c01618.

[3]

D. Pereira, M. Sardo, R. Vieira, I. Marín-Montesinos, and L. Mafra, "Enhancing CO₂ capture via fast microwave-assisted synthesis of the calf-20 metal-organic framework," *Inorganic Chemistry*, vol. 64, no. 7, pp. 3302–3310, Feb. 2025, doi:10.1021/acs.inorgchem.4c04727

[4]

A. M. Najafi, S. Soltanali, F. Khorashe, and H. Ghassabzadeh, "Effect of binder on CO₂, CH₄, and N₂ adsorption behavior, structural properties, and diffusion coefficients on extruded zeolite 13X," *Chemosphere*, vol. 324, p. 138275, May 2023, doi:10.1016/j.chemosphere.2023.138275

[5]

J.-B. Lin *et al.*, "A scalable metal-organic framework as a durable physisorbent for carbon dioxide capture," *Science*, vol. 374, no. 6574, pp. 1464–1469, Dec. 2021, doi:10.1126/science.abi7281

[6]

M. L. Yeboah, X. Li, and S. Zhou, "Facile fabrication of biochar from Palm Kernel Shell Waste and its novel application to magnesium-based materials for hydrogen storage," *Materials*, vol. 13, no. 3, p. 625, Jan. 2020, doi:10.3390/ma13030625

[7]

S. Sircar, "Basic research needs for design of adsorptive gas separation processes," *Industrial & Engineering Chemistry Research*, vol. 45, no. 16, pp. 5435–5448, Feb. 2006, doi: 10.1021/ie051056a.

[8]

Y. Wei *et al.*, "Efficient Xe selective separation from Xe/KR/n₂ Mixtures over a microporous calf-20 framework," *RSC Advances*, vol. 12, no. 28, pp. 18224–18231, 2022, doi:10.1039/d2ra02768b

[9]

B. Bensafi, N. Chouat, and F. Djafri, "High performance of carbon dioxide adsorption of mesoporous mordenite synthesized in the presence of N,N-dimethylaniline - research on chemical intermediates," SpringerLink, <https://link.springer.com/article/10.1007/s11164-017-3085-2> (accessed Dec. 10, 2025).

[10]

M. S. Khodadadi and A. Ghaemi, "Predicting flue gas component adsorption in calf-20: A molecular dynamics study anchored by experimental CO₂ Capture," *Chemical Engineering Journal Advances*, vol. 24, p. 100931, Nov.

[11]

M. B. Pourazar, T. Mohammadi, M. R. Jafari Nasr, M. Javanbakht, and O. Bakhtiari, "Preparation of 13x zeolite powder and membrane: Investigation of synthesis parameters impacts using experimental design," *Materials Research Express*, vol. 7, no. 3, p. 035004, Mar. 2020. doi:10.1088/2053-1591/ab7217
<https://iopscience.iop.org/article/10.1088/2053-1591/ab7217>

[12]

T. T. T. Nguyen, B. M. Balasubramaniam, N. Fylstra, R. P. S. Huynh, G. K. H. Shimizu, and A. Rajendran, "Competitive CO₂/H₂O adsorption on CALF-20," *Industrial & Engineering Chemistry Research*, vol. 63, no. 7, pp. 3265–3281, Feb. 2024, doi: 10.1021/acs.iecr.3c04266.

[13]

J. Wang, L. A. Stevens, T. C. Drage, and J. Wood, "Preparation and CO₂ adsorption of amine modified Mg–AL LDH via Exfoliation Route," *Chemical Engineering Science*, vol. 68, no. 1, pp. 424–431, Jan. 2012. doi:10.1016/j.ces.2011.09.052

Investigating The Effects of Protecting Groups on the Thermodynamics and Kinetics of Alanine

Mohamad Adan, Ahmad Adam B¹ and Ahmad Rafiq, Aliff M B²

¹Department of Chemical Engineering, Imperial College London, UK

This manuscript was compiled on December 11, 2025

L-alanine is a prototypical amino acid active pharmaceutical ingredient (API) precursor that is closely linked to sugar metabolism, energy production for muscle tissue and immunity boost. Tight quality specifications are imposed on APIs to ensure efficacy and safety, whereas crystallisation will facilitate improving yield and purity. In the synthesis route of peptides, its protected derivatives are common process-related impurities. Hence, this study focuses on the effects of the protecting groups on the thermodynamics (solubility) and kinetics of alanine (i.e. induction time). The effects of impurities on solubility of the main component were also investigated. It was found that the protecting groups N-tert-butoxycarbonyl (Boc) and benzyloxycarbonyl (Cbz) reduces the solubility by a factor of 4 and 13, respectively. No clear correlation between protecting groups and induction time was shown due to the stochastic nature of the phenomena. It was also discovered that the presence of Cbz increases the pre-exponential factor, A and thermodynamic factor, B of alanine. As for the impurity experiment, the solubility of the main component at 25°C was found to be insensitive to the presence of the impurity. However, in a solution where alanine is the dominant species and Cbz-alanine is present as the limiting component, Cbz-alanine becomes 2-5 fold more soluble than itself in deionised water at different temperatures. This will become pertinent in the synthesis route of alanine, especially during washing and crystallisation.

Alanine | Protecting Groups | Solubility | Induction Time | Crystallization | Pre-Exponential Factor | Thermodynamical Factor | Impurities

Introduction

An Active Pharmaceutical Ingredient (API) is a chemical substance that acts as the major contributor to therapeutic effects for medicinal uses. Due to its stringent purity levels mandated for pharmaceutical uses, the favoured method to synthesise operates under heterogeneous solid-liquid conditions, which facilitates efficient separation and purification of the target compound [18]. In general, impurities exceeding 0.1 wt% must be acknowledged and quantified during synthesis to ensure safety and efficacy [6]. Pharmaceutical products are prone to impure substances depending on their synthesis routes during formulation and ageing of drugs. Hence, an ultra-high level of purity is crucial in drug manufacture as lives and safety are an issue of concern [4]. In the context of alanine, its protected derivatives are often found as impurities as they are used in the synthesis route of the target molecule.

Amino acids have been shown to play a critical role in physiological regulatory functions, energy production, neurotransmitters in the central nervous system as well as for metabolic processes. Alanine specifically has many uses; one of its notable functions acts as a promoter of poly-ribosomal aggregation and plays an important role in hepatic protein synthesis [20]. In a recent study, L-alanine has shown importance in anti-proliferative potential against human lung adenocarcinoma and mammary gland adenocarcinoma [21]. Poly-alanine peptide chains also boast a strong contribution in the pharmaceutical field due to their importance and varied uses. *Pa*-MAP 1.9 is a poly-alanine-rich molecule acting as a rationally designed antimicrobial peptide that shows promising results to act as a key ingredient to fight against pathogenic bacteria infections [3]. Figure 1 shows a microscopic image of alanine under 10 X magnification.

The traditional approach to synthesising an alanine peptide chain is solution-phase peptide synthesis, where both amino acids and intermediates are dissolved in an aqueous solution,

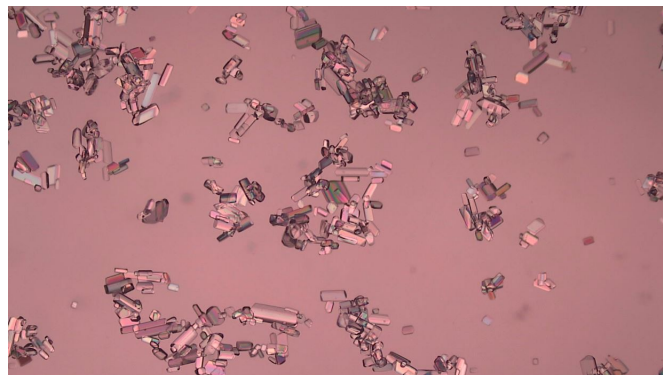


Fig. 1. Microscopic picture of alanine crystals in DI water at 10 X magnification

and a cycle of coupling and deprotection generates the desired peptide chain. Purification follows after the peptide is isolated for each cycle, which makes this step the bottleneck of the process. Apart from being time-consuming, the rapid changes in the peptide's environment increase the risk of product loss, subsequently decreasing the yield percentage[23].

A more recent approach makes use of a solid phase which acts as a support for the amino acids. The Solid Phase Peptide Synthesis (SPPS) is a stepwise formation of peptide bonds between amino acids attached to a hydroxyl or amino resin (solid phase). SPPS of Alanine employs two main strategies: either using Boc/Bzl or the Fmoc/tBu. Boc-alanine and Fmoc-alanine are peptide chains with protecting groups consisting of N-tert-butoxycarbonyl-L-alanine and N-9-Fluorenylmethoxycarbonyl, respectively. The N-protected C-terminal amino acid residue is attached to the solid phase, resulting in a C-terminal acid or C-terminal amide peptide. Following this first amino loading, peptide sequences are then linearly arranged from C-terminus to N-terminus. Subsequent iterative steps of de-protection, activation, and coupling are performed, as in the classic method, until

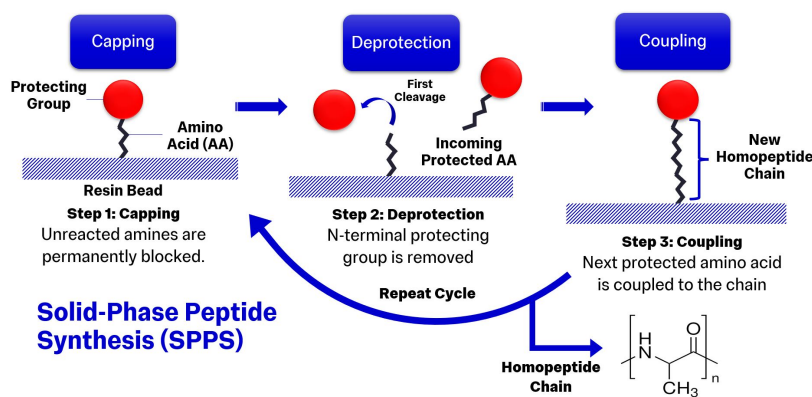


Fig. 2. Visualisation of the solid-phase peptide synthesis of a long homo-alanine chain

the peptide attains the specified chain length. This method is revolutionary as the insoluble resin allows for direct washing of unreacted reagents after each step, preventing degradation of the target molecule and is less time-consuming [1].

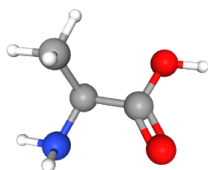


Fig. 3. 3D structure of alanine molecule [15]

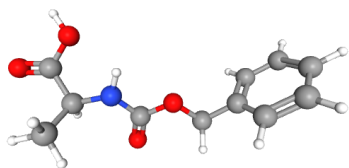


Fig. 4. 3D structure of Cbz-alanine molecule [17]

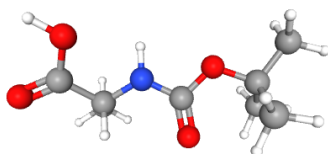


Fig. 5. 3D structure of Boc-alanine molecule [16]

Both methods (solution-phase and solid-phase peptide synthesis) demand a thorough understanding of the component solubility and crystallisation thermodynamics. We have executed this study under laboratory conditions, and while scale-up typically widens the gap between laboratory yields and industrial outputs, a solid grasp of the fundamentals underpins subsequent process optimisation. A critical framework is mapping the solubility-temperature profile of alanine and its protected derivatives i.e Cbz-alanine and Boc-alanine in deionised water to elucidate their underlying thermodynamic behaviours. Equally critical would also be to define the induction time of alanine and its protected derivatives under

different supersaturation ratios. This will allow us to analyse the effects of the protecting groups on alanine and compare that to alanine. After establishing an understanding of the kinetics and thermodynamics of the pure components, we can explore how the impurities affect the solubility of alanine. This can then be translated into industrial applications where one can adjust their synthesis route to accommodate the behaviours of aforementioned impurities.

Background

A. Classical Nucleation Theory. Crystals, by definition, are the most highly ordered particulate of a non-living material characterised by an orderly 3D array, consisting of atoms, ions or molecules [7]. The process of forming said particle is called crystallisation. Crystal dissolves in a solvent and depends on the temperature and pressure of the system. Solubility is defined as the amount of solvent able to dissolve in a solution. When the maximum amount of solute is dissolved in a solvent where it is thermodynamically stable, it results in a phase equilibrium defined as a saturated solution. However, when more solute is present in the system than at saturated conditions, the system is considered to be supersaturated [14]. A supersaturated solution can be achieved by evaporation or cooling, subsequently changing the concentration and temperature of the system, respectively. The degree of supersaturation can be calculated with Equation 1:

$$S = \frac{c}{c^*} \quad [1]$$

where S is the degree of supersaturation, c is the concentration of the current system, and c^* is the concentration of the solution at saturated conditions. For a supersaturated system, $S > 1$, whereas for a saturated solution, $S = 1$. The solubility of the particle at size r can be described to be proportional to the exponential of $\frac{1}{r}$, as demonstrated in the Ostwald-Freundlich equation, 2 [5].

$$\ln S = \frac{2M\gamma_{int}}{n_i RT \rho_s r} \quad [2]$$

where ρ_s is the density of the particle, n_i is the number of ions formed in the solution upon dissociation, M is the molecular weight and γ_{int} is the interfacial surface energy.

There are two routes of nucleation: primary and secondary. For primary nucleation, crystal presence occurs with no initial seeding or provision of nuclei. Nucleation may occur in a

homogeneous system, where crystal formation is driven by supersaturation only, or in a heterogeneous system, where the presence of insoluble material aids the formation of crystals. For secondary nucleation, crystals are formed after some initial seeding of the same-species nuclei is applied to the system, aiding in crystal formation. Formation of crystals is often governed by the classical nucleation theory. This theory works under the assumption that a crystal is perfectly spherical and that interfacial surface energy, γ_{int} , is isotropic. The theory relates the Gibbs free energy to that of the size of the crystal [10].

$$\Delta G_r = -\frac{4}{3}\pi r^3 \Delta G_V + 4\pi r^2 \gamma_{\text{int}} \quad [3]$$

where ΔG is the Gibbs free energy (J). This equation shows that there is a trade-off present when a crystal is formed, as when the crystal grows, the volume term is more favourable than the surface term. A critical radius, r_c , must be achieved by the nucleus for crystals to appear. The crystal radius can be found by taking the derivative of Equation 3 to obtain Equation 4 and Equation 5

$$0 = -4\pi r_c^2 \Delta G_V + 8\pi r_c \gamma_{\text{int}} \quad [4]$$

$$r_c = \frac{2\gamma_{\text{int}}}{\Delta G_V} \quad [5]$$

Substituting Equation 5 into Equation 3 will result in the following:

$$\Delta G_c = \frac{4}{3}\pi r_c^2 \gamma_{\text{int}} \quad [6]$$

The thermodynamics of crystal formation can be related to the kinetics of nucleation. Using the Arrhenius equation to describe the rate of a reaction, as described in Equation 7

$$J = A \exp -\frac{\Delta G}{KT} \quad [7]$$

where J is the nucleation rate and A is the pre exponential factor. By first rearranging Equation 2, then substituting in Equation 6, we obtain

$$r^2 = \left(\frac{2M\gamma_{\text{int}}}{n_i RT \rho_s \ln S}\right)^2 \quad [8]$$

$$\Delta G_c = \frac{4}{3}\pi \left(\frac{2M\gamma_{\text{int}}}{n_i RT \rho_s \ln S}\right)^2 \gamma_{\text{int}} \quad [9]$$

Finally, we substitute 9 into Equation 7 to obtain an expression of nucleation rates in terms of solubility parameters.

$$J = A \exp -\frac{16\pi\gamma_{\text{int}}^3\nu^2}{K^3T^3(\ln S)^2} \quad [10]$$

where ν is the specific volume of the molecule. We arrive at an equation dispensable to extrapolate and predict the pre-exponential factor and interfacial surface energy of a crystal. In this study, we are interested in observing how the protected groups present in Cbz-alanine and Boc-alanine affect the following parameters. Experimentally, these values can be obtained by varying the supersaturation ratio and observing how the nucleation rate changes.

B. Induction Time. Induction time is a stochastic process due to several factors, such as molecule interactions and variation within temperature differences along a vessel. Induction time is then expressed as a function of the probability that a N number of nuclei are found. This results in a Poisson Distribution [25].

$$P_t = 1 - \exp N \quad [11]$$

According to Shiau [22], nucleation rates can be measured by observing the induction time of the crystal. Induction time is defined as the delay between the attainment of a supersaturated solution and the onset of newly formed crystal detection. In general, the induction time is inversely proportional to the nucleation rate. If we take the induction time to be the onset of the first nucleus at time t_i , this corresponds to a number of nuclei $N(t) = 1$.

$$N(t) = V \int J(t) dt \quad [12]$$

$$1 = V J t_i \quad [13]$$

where V is the volume of solution and t_i is the induction time. Hence, introducing Equation 13 with Equation 10, we obtain the following

$$\ln t_i = -\ln A + \frac{B}{(\ln S)^2} \quad [14]$$

where $B = \frac{16\pi\gamma_{\text{int}}^3\nu^2}{K^3T^3}$. Therefore, by varying the supersaturation ratio $\ln S$ and plotting it against the induction time t_i , we are able to plot a linear graph with the y-intercept being the pre-exponential factor $\ln A$ and the gradient being the constant B . According to Brandel [2], in a heterogeneous system, the thermodynamic factor B is a balance between 3 different interfacial energies: that between nucleus and heterogeneous particle, nucleus and the solution and the solution with the heterogeneous particle. The pre-exponential factor describes the statistical process which a crystal unit attaches itself to a nucleus to reach the critical radius. Further discussion will be made in Section J.

C. Effects of Impurities. Impurities are able to significantly affect the crystallisation process of pharmaceutical products. The solubility of paracetamol is significantly reduced in the presence of impurity 4-nitro-phenol and 4-chloroacetanilide, along with its kinetic parameter inhibited [9]. According to Lia [12], impurities will severely affect safety, efficacy and quality aspects of APIs. Impurities arise from various routes, including oxidation, substitution, isomerisation, deamidation, racemisation, etc. In [11], an investigation on the growth kinetics of crystals from impure solution mixtures were performed and was able to find that supersaturation, temperature and impurity concentration played a significant role in the impurity action. In this work, we have looked into how impurities will affect the solubility of a compound, subsequently affecting crystallisation and the quality of the product. In [24], it was found that the presence of acetanilide decreased the total product recovery by up to 15 %. There are clear relationships between impurities and how they influence the final product; hence, this becomes a benchmark of our study on how impurities affect the solubility of alanine, Boc-alanine and Cbz-alanine.

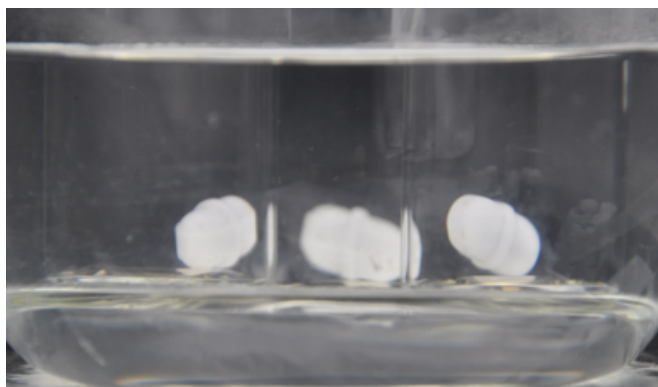


Fig. 6. Photo taken of induction time experiment before nucleation occurs

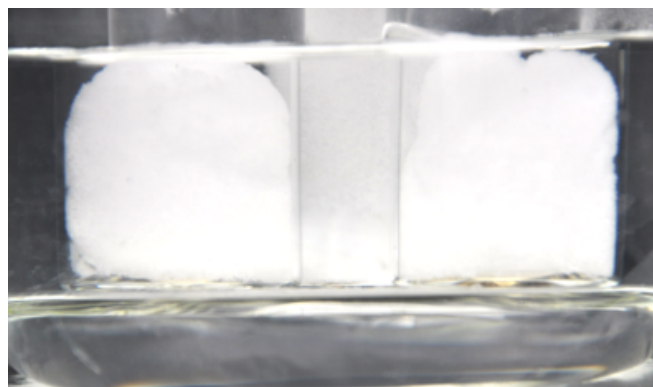


Fig. 7. Photo taken of induction time experiment after nucleation occurs

Materials and Methodology

D. Materials. Alanine (L-alanine; $\geq 99\%$ purity) was obtained by Thermo Scientific Chemicals, Cbz-alanine (Carbobenzoxyl-L-alanine; $\geq 99.83\%$ purity) and Boc-alanine (tert-Butyloxycarbonyl-L-alanine; $\geq 99\%$ purity) were procured from BLDpharm.

Solution preparation. Fisherbrand 5 and 10 ml Plastic Syringe, Starlab $0.45\ \mu\text{m}$ Sterile Syringe Filter, BD Blunt Fill Needle ($1.2\ \text{mm} \times 40\ \text{mm}$), VWR Weighing Paper and Fisherbrand 15 ml test tubes were used in preparing the solution. The test tubes are placed in the Dlab HCM100-Pro Thermoshaker.

HPLC. Methanol ($\geq 99\%$ purity), Ammonia Formate ($\geq 99\%$ purity), trifluoroacetic acid (TFA; $\geq 99\%$ purity) were used as mobile and stationary phase respectively. A Supelco 2ml clear glass vial ($12 \times 32\ \text{mm}$) was used as the vessel.

Induction time. Nikon DX 18-105 mm camera was used to capture the photos. A Huber water bath was used to control the temperature of the water.

E. Saturated Solution Preparation. Alanine, Boc-alanine and Cbz-alanine crystals are fed into 15ml test tubes until an excess amount is present. Deionised water is then added to the tubes and shaken until all solids are in contact with the solvent. The tubes are then left in the thermoshaker running at 25°C at 800 rpm for over 4 hours. This procedure is repeated for temperatures 30°C , 35°C , 40°C and 45°C . It is vital to check the tubes periodically to ensure excess undissolved solids are still present within the solution. Otherwise, more solute is added until no more solids can be dissolved. During solution extraction, a 5 ml syringe is used to collect the liquid content of the tube. A $0.45\ \mu\text{m}$ hydrophobic filter is used to separate any unfiltered undissolved solids. The filtered solution is collected in labelled vials.

F. HPLC Method. A calibration curve is prepared by preparing solutions of alanine, Boc-alanine and Cbz-alanine at known concentrations i.e 5wt%, 10wt%, 15wt%, 20wt%, 25wt%, 30wt%. The aforementioned samples are left to settle, and $100\ \mu\text{l}$ of the saturated solution is collected via syringe and filtered into a 2 ml vial. $900\ \mu\text{l}$ of deionised water is added into the small vial to dilute the solution. The mobile phase used is 5% methanol in 95% DI water with 10 mM ammonia formate at pH 2.6 Trifluoroacetic acid (TFA) (method provided by the student

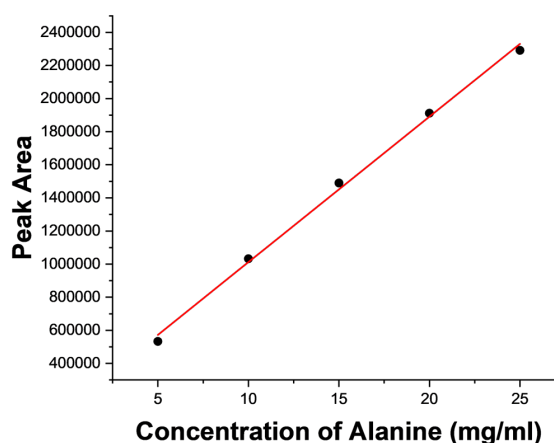


Fig. 8. Calibration curve for known concentration of alanine in DI water using high-performance liquid chromatography

supervisor). The calibration curve is used to determine the concentration of our solute in further experiments. This is done by fitting a best-fit line to the curve and determining the equation parameters. This equation can then be used to back-calculate the solubility of alanine in an unknown concentration. Figure 8 is an example of the calibration curve done on alanine.

G. Induction Time Experiment. 30 ml of saturated solution is collected by syringe and is filtered into a 30 ml vial carrying a magnetic stirrer. It is important that the addition of the solution occurs after the magnetic stirrer is added to avoid heterogeneous nucleation from occurring. The vial is closed off and sealed with a parafilm tape. This vial is placed into a water bath running at 20°C . A camera is set in a well-lit and covered area and is automated to capture pictures at a 60-second interval. The magnetic stirrer is set to run at 250 rpm. The setup is left to run for an hour before checking if any samples have nucleated. Stop the experiment once all samples have nucleated. The picture with nucleated samples is taken as a reference, and the timestamp of the picture recorded. The time difference between the picture taken and the start of the experiment is calculated and compiled. This experimental setup is repeated for every sample prepared. For recycled samples, the nucleated samples are sent to a water bath at 40°C until crystals are redissolved. This sample is then reintroduced

to the 20°C water bath. Figure 6 and 7 depict the setup of the experiment. Considering that the induction time of a compound is highly stochastic, it is imperative that the number of repeats to account for outliers and random events such as non-nucleation or extremely long induction time. By increasing the number of data points, a distribution graph can be obtained, from which the median and the average can be obtained.

Discussion and Results

H. Solubility-Temperature Profile. We were able to find that alanine and its derivatives increase in solubility as the temperature increases. Dissolution of a crystal is generally endothermic, absorbing heat in the process. Using Le Chatelier's principle, an increase in heat will shift the equilibrium to favour dissolution, and hence the amount of solute dissolved in solvent increases [13]. Despite showing similar trends, the presence of the protecting molecule changes the overall solubility of the compound. As seen in Table 1, at 25°C, the solubility of alanine is 141 ± 4 mg/ml. With the addition of the benzyloxycarbonyl (Z) group to the alanine molecule, the solubility of the compound is reduced by a factor of 13, making the solubility of Cbz-alanine to be 10 ± 1 mg/ml. The addition of a tert-butoxycarbonyl reduces its solubility by a factor of around 4, making the solubility 35 ± 2 mg/ml.

Table 1. Solubility data for various compounds at 25°C

Compound	Solubility (mg/ml)	Percentage Error (%)
Alanine	141 ± 4	2.8
Cbz-alanine	10 ± 1	10
Boc-alanine	35 ± 2	5.7

H.1. Steric hindrance. Tert-butoxycarbonyl (Boc group) and benzyloxycarbonyl (Cbz group) are bulky in nature and bigger relative to the alanine molecule. The large presence of the molecule reduces the likelihood of hydrogen bond formation, subsequently reducing its overall solubility in deionised water.

H.2. Non-polarity. Benzyloxycarbonyl (Z) is considered a large non-polar molecule. The addition of this molecule subsequently reduced the overall polarity of alanine, further reducing the likelihood of hydrogen bond formation to the molecule. This also explains why Cbz-alanine has a much lower solubility than Boc-alanine due to this added restricting factor.

Although alanine, Boc-alanine and Cbz-alanine show a positive increase as temperature increases, the rate at which they increase, $\frac{\delta S}{\delta T}$ is different due to the varying gradients in their profiles. Figure 9 compares the difference between the profiles. Hence, for the same temperature difference, ΔT , different supersaturation ratios are achieved, subsequently affecting other key parameters such as induction time.

I. Induction Time. The degree of saturation is calculated using Equation 1. The supersaturation recorded in Table 2 is the saturation of the solution recorded at 30°C, c , compared to the saturation of the solution at 25°C, c^* . From our experiments, we were able to deduce that the induction time of alanine Cbz-alanine is 94 ± 24 and 24 ± 11 minutes respectively. For Boc-alanine, samples did not appear to nucleate after leaving them to nucleate for over 24 hours. A summary of this information is displayed in Table 2. There were no clear trends on how the

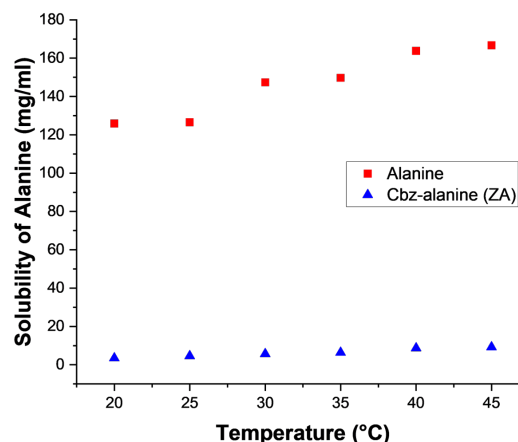


Fig. 9. HPLC results of solubility-temperature profile for alanine, Boc-alanine and Cbz-alanine in deionised water at 25°C under room conditions

protecting groups affected the induction time of alanine. This is due to a few factors:

I.1. Stochasticity. Induction time is highly stochastic and is dependent on various external factors. By increasing the number of data points, outliers such as non-nucleation or extremely long nucleation time (longer than a day) are excluded in our calculation to find the mean of induction time. Despite that, this does not discount the stochastic nature of the induction time's effect on the actual average time of our compounds. Hence, a larger number of repeats is needed to obtain a value with a lower percentage error.

I.2. Solubility-temperature change, $\frac{\delta S}{\delta T}$. The experiment was run under the same temperature difference, with the solution being prepared at 30°C and was left to nucleate at 25°C. However, comparing the solubility-temperature profiles of the different compounds mentioned in section H, at different temperature differences results in different solubility for the three compounds. A higher supersaturation ratio would lead to a lower induction time as a larger thermodynamic force is present. As for Cbz-alanine, the gradient for a temperature change of $\delta T = 30^\circ\text{C} - 25^\circ\text{C}$ is steeper than that of alanine. Hence, Cbz-alanine is more likely to induce quicker at $\delta T = 30^\circ\text{C} - 25^\circ\text{C}$ than alanine due to its higher supersaturation ratio. Interestingly, the same cannot be said for Boc-alanine. Despite having a higher supersaturation ratio than alanine under the same conditions, Boc-alanine refused to show signs of nucleation under 24 hours. This gives rise to other external factors involved, such as, but not limited to, ionic strength, specific ion effects and interfacial surface energy of the compound. Due to the susceptibility of induction time to be influenced by various external factors, it is deemed necessary to repeat this experiment with a higher number of repeats. This subtopic has been further discussed in the Supplementary Data Section, L, where the number of samples is increased by reheating solutions to produce undersaturated solutions at higher temperatures.

J. Pre-Exponential Factor, A and Thermodynamic Factor, B . To relate solubility and induction time of our components, we employed Equation 14, to obtain a linear relationship. The value for supersaturation ratio, S is calculated using Equation

Table 2. Induction time data for various compounds 30°C - 25°C

Compound	Induction Time (min)	Supersaturation Ratio
Alanine	97 ± 24	1.2
Cbz-alanine	24 ± 11	1.5
Boc-alanine	≥ 1440	1.3

1. Then, the values for pre-exponential factor, A and thermodynamic factor, B can be interpolated. Due to Boc-alanine's long induction time, this analysis was only done for alanine and Cbz-alanine.

It is concluded that from the results of this experiment (Table 3), the presence of the benzyloxycarbonyl (Cbz) group increases both the pre-exponential factor and thermodynamic factor of alanine. As seen in Figure 10 and 11, the error bars are significant and deviate largely. This is the case due to the stochastic nature previously mentioned in Section I. To mitigate this issue, more repeats need to be made to produce an average induction time with a higher confidence interval.

Table 3. Pre-exponential factor, A and thermodynamic factor, B of alanine and Cbz-alanine

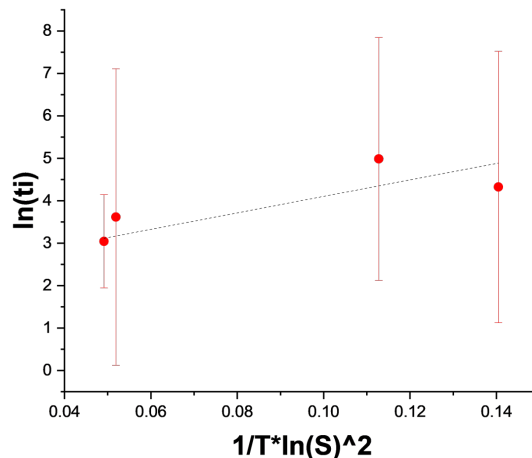
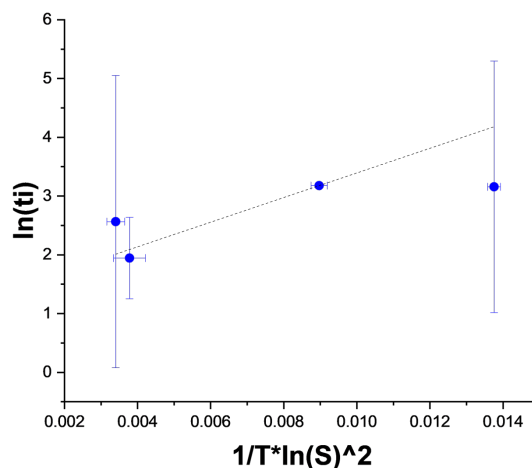
Factor	Alanine	Cbz-alanine
Pre-Exponential Factor, A	0.06141	0.13642
Thermodynamic Factor, B	13.88	196.24
R^2 Value -	0.6629	0.6503

J.1. Pre-exponential Factor, A . As discussed in Section B, the pre-exponential factor, A , is a statistical description of how likely a building unit of crystal attaches itself to a nucleus until it reaches a critical radius. This critical radius is necessary to overcome for the crystal to form. This was previously discussed in Equation 5. The pre-exponential factor is a product of the concentration of nucleation sites, C_0 , the attachment frequency, f^* , and the Zeldovich factor, z [2].

$$A = z f^* C_0 \quad [15]$$

The Zeldovich factor, z is a probabilistic description of a cluster larger than the nucleus to decay or grow. The attachment frequency, f^* , determines how many nuclei increase in size. In the context of alanine, the higher value of A mathematically describes that a crystal is more likely to nucleate. In this study, we found that with the presence of the benzyloxycarbonyl group, the pre-exponential factor increases by a factor of 1.8. The addition of the protecting group assists in the likelihood of building units to attach to the nucleus. With the presence of the bulky Cbz-group, the surface area of the molecule is larger and allows for better interaction between building units and nuclei, further increasing and improving the frequency of successful attachment onto the nuclei.

J.2. Thermodynamical Factor, B . As for the thermodynamic factor, B , it represents the creation and reduction of interfacial energy between the interaction of nuclei in the solution. B is a function of interfacial energy, γ_{int} , as seen in Equation 14. In the case for homogeneous nucleation that ran in this experiment, the interfacial energy involved is between the solid crystal and the liquid solution interface. An energy barrier is needed to


Fig. 10. Natural logarithm of induction time of alanine against the inverse of supersaturation ratio squared

Fig. 11. Natural logarithm of induction time of Cbz-alanine against the inverse of supersaturation ratio squared

overcome for successful growth to occur. Therefore, larger values of γ_{int} lead to larger thermodynamical factor, B . This would mean that the molecule has a higher energy barrier to overcome to successfully create a solid interface in the solution. In the case of alanine, we observe that the presence of benzyloxycarbonyl (Cbz) group increases the thermodynamic factor, B . The presence of the bulky protecting group creates a higher energy barrier as more solid-liquid interface is in contact due to a higher surface area. This subsequently increases the energy barrier needed for Cbz-alanine to crystallise in the solution.

In this study, there is evidence on the effects of the protecting group on the solubility and the induction time of the molecule. Specifically for Cbz-alanine, the addition of the protecting group increases the surface area of the molecule, improving the success rate of building units to attach to the nuclei. However, this comes with a trade-off as the presence of the bulky protecting group increases the energy barrier for the crystal to surpass for successful crystal growth. However, this conclusion must

first be checked with a larger number of repeats to remove outliers in induction time to produce better results. Hence, this conclusion can only be done in a qualitative sense, rather than quantitative.

K. Effects of Impurities on Solubility. To investigate the relationship of impurities present and their effects towards solubility, we explored the solubility of pure alanine, Boc-alanine and Cbz-alanine in the presence of different wt% of the other derivatives.

The wt% of an impurity is calculated as the ratio of the mass of impurity, m_I , over the total mass of the system, m_T . The total mass of the system consists of the mass of the impurity, the mass of alanine present in the solvent at temperature T , m_A , and the mass of solvent used, m_{solv} .

$$m_T = m_I + m_A + m_{solv} \quad [16]$$

$$wt\% = \frac{m_I}{m_T} \times 100\% \quad [17]$$

Hence, by rearrange Equation 17, we obtain the amount of impurity added into the saturated system.

$$m_I = m_s \left(\frac{wt\% \times 0.01}{1 - (wt\% \times 0.01)} \right) \quad [18]$$

We observed how impurity Boc-alanine and Cbz-alanine affected the solubility of pure alanine, and how impurity alanine affected the solubility of pure Cbz-alanine and pure Boc-alanine. Strikingly, the addition of impurity Boc-alanine or Cbz-alanine does not significantly affect the solubility of pure alanine at 25°C. Although the solubility fluctuates between different wt% of impurity added, no clear trend can be observed, and therefore, the presence of alanine derivatives impurities does not produce statistically significant changes at 25°C.

It is also shown that impurity alanine does not affect the solubility of Boc-alanine and Cbz-alanine at 25°C. At higher addition of impurity ($\geq 50\%$) was decided unnecessary as the volume of added impurity would overwhelm the solution, severely decreasing the mass transfer of the solute to solvent and therefore render the experimental result flawed. Impurity in the presence above 50% would also not be considered an impurity by our standard definitions.

K.1. Cbz-alanine in alanine solution. Interestingly, we were able to discover that the solubility of Cbz-alanine is 2-5 times higher in alanine solution. We were able to obtain a similar result at differing temperatures and at differing alanine concentrations in the solution. This was not observed for Boc-alanine in a solution of alanine. Therefore, it is concluded that the solubility of Cbz-alanine is 2-5 times in DI water + alanine than in DI water alone, therefore supporting that the alanine molecule promotes the dissolution of Cbz-alanine. Several reasons can be attributed to this phenomenon. In a solution of alanine + DI water, the polarity would increase, further assisting in hydrogen bond formation on the Cbz-alanine molecule. Due to the zwitterionic nature of alanine, co-solvation can also occur, where the ammonium ion NH_3^+ electrostatically associates with the deprotonated Cbz-alanine carboxylate. This forms a transient ion pair and H-bonded complexes. This could prevent the disruption of hydrophobic aggregation of the benzyloxycarbonyl group, subsequently increasing in solubility.

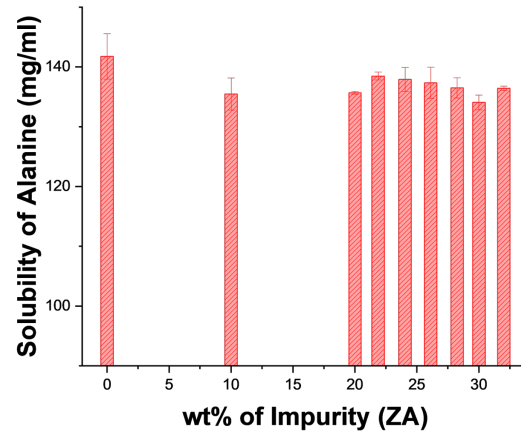


Fig. 12. Solubility of alanine solution with different weight percentages of Cbz-alanine impurity added into the solution

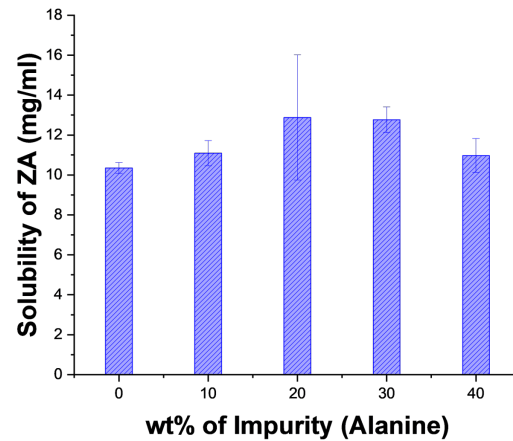


Fig. 13. Solubility of Cbz-alanine solution with different weight percentages of Cbz-alanine impurity added into the solution

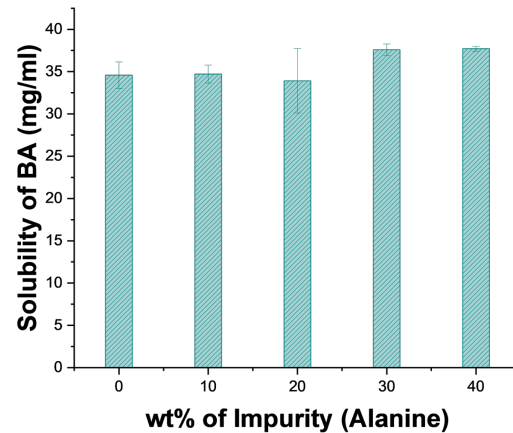


Fig. 14. Solubility of Boc-alanine solution with different weight percentages of Cbz-alanine impurity added into the solution

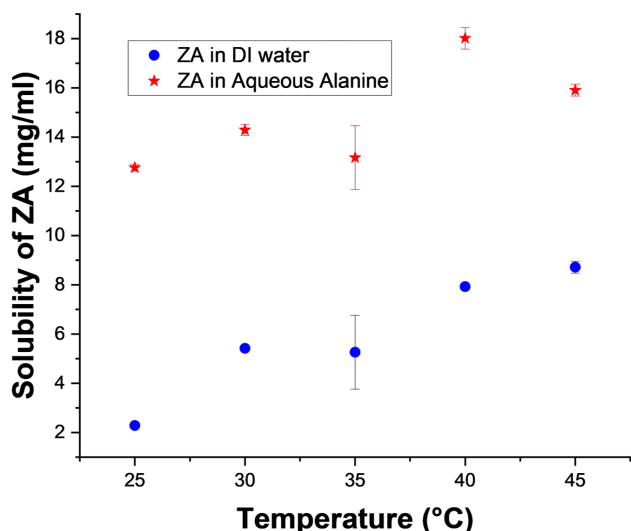


Fig. 15. Solubility of Cbz-alanine in DI water and alanine solution at temperatures between 25°C - 45°C

Supplementary Data

L. Effects of Recycling with Induction Time. As proven earlier, induction time is a stochastic phenomenon and is severely affected by external factors such as impurities, agitation and shape of vessels. We explored the effects of recycling a sample where crystals have already been induced. After preparing a saturated solution of alanine at 25°C, 3ml of the solution is transferred to a vial and sealed. The solution is transferred to a water bath running at 20°C and the time taken for the first crystal to form is recorded. Then, the already-induced sample is reheated to a high temperature such that all previously induced crystals are redissolved. Subsequently, the vial is weighed and re-submerged in the water bath at 20°C. The induction time for the 1st recycle is recorded. This process is repeated for n number of recycles. By comparing the induction time values of recycled components in relation to the fresh sample, we can define a percentage difference in Equation 19

$$\Delta\% = \frac{|R_0 - R_n|}{R_0} \times 100\% \quad [19]$$

By plotting the percentage difference with the number of recycles the sample has undergone, we can determine any trends present, as done in 16. The percentage difference increases and decreases randomly across different vials, confirming that the stochastic nature of induction time is maintained. This gives rise to possible outlooks where the induction time of a compound can be experimentally determined by recycling the samples, instead of making fresh samples, given that the supersaturation ratio is kept constant (by ensuring no sample is lost due to spills or evaporation). This will reduce the burden of making new samples in large quantities and promote sustainable practices by avoiding unnecessary waste.

M. Effects of Agitation and Settling Time on Solubility. Agitation and settling time are very important parameters that affect the overall solubility. By altering these parameters, the 'apparent' solubility will present itself to be higher or lower than what is thermodynamically and kinetically possible.

During dissolution, there is a mixed dependence between crystal interfacial kinetics and mass transfer. However, at

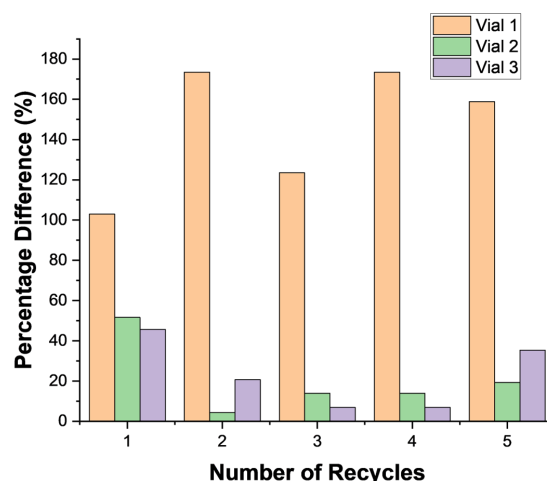


Fig. 16. The percentage difference of induction time of the n^{th} number of recycles to its fresh sample

higher undersaturations, the dissolution becomes more limited by mass transfer [19]. This creates a boundary layer between the crystal and the solution, which can be further reduced with an increase in hydrodynamic disturbances i.e. agitation. This is further supported with mass transfer coefficient k_m being a function of Reynold's number Equation 21 explains the relationship between mass transfer and fluid velocity

$$Re = \frac{\rho U D}{\mu} \quad [20]$$

$$k_m \propto Re^a Sc^b \quad [21]$$

In our results, we have observed that the 'apparent' solubility of alanine is halved when the shaker is changed to 500 RPM instead of 800 RPM. Due to the low agitation, not all solids are in contact with the solution. Under low rotation per minute, the vessel has not been well shaken, so that the dry solids are periodically exposed to the solution, subsequently becoming wet. Rather, the dry solids are protected by the wet solid formed above them. Hence, the supposed amount of solids that can be accommodated in the solution is low, as its capacity has not been fully utilised, producing an undersaturated solution despite leaving it shaken for longer periods.

On the other hand, a solution can be in its metastable state as more crystals have dissolved in the solvent than what is thermodynamically stable. This transient stage arises due to the stochastic nature of the induction time. During dissolution, crystal formations are broken down into their substituents while forming hydrogen bonds. This step increases the solubility of the alanine momentarily as more crystals are broken down into its ionic components, depending on the surrounding temperatures, molecular orientation and other external factors. When this solution is allowed to settle, crystals will nucleate such that the saturation level is thermodynamically stable. However, if the solution is given less time to settle, the likelihood that the system is in its metastable state is higher, leading to higher values of solubility than what is thermodynamically feasible [8].

In this experiment, the solubility of alanine was reported to be 1.25 times higher than its expected solubility. The values

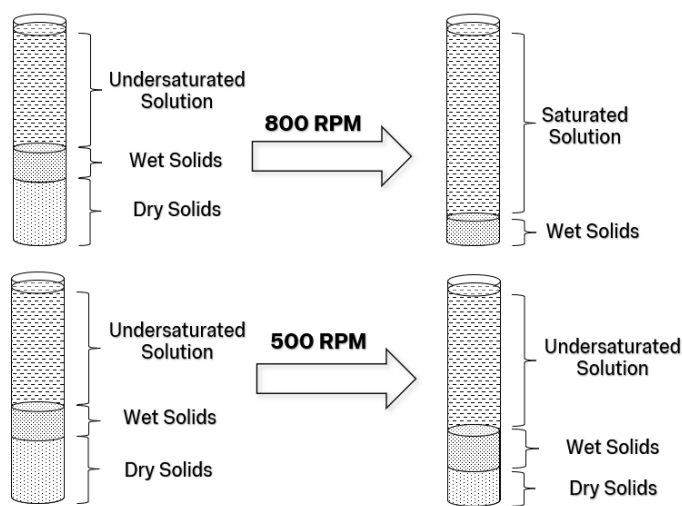


Fig. 17. Visualisation of the effects on agitation on mass transfer

reported have high standard deviations, as the likelihood of the induction occurring during measurement is high. Hence, it is very imperative that the solution is given ample time to settle at the experimental conditions for more than 4 hours to reduce the likelihood of induction occurring during measurement.

Conclusion

In this study, we have concluded that the solubility of alanine, Boc-alanine and Cbz-alanine increases with temperature, with no retrograde behavior detected. With the presence of the tert-butoxycarbonyl group, the solubility of alanine is reduced by a factor of 4 due to steric hindrance whereas benzyloxycarbonyl reduces the solubility by a factor of 14 due to steric hindrance and hydrophobicity. This plays an important role during SPPS as chemical impurities can be filtered easily without the worry of dissolution into the solvent. This investigation was, however, restricted between 20°C and 45°C. These reported trends may be different and would subsequently affect our current conclusion.

As for induction time, protecting groups have no clear effects on alanine, as the process is highly stochastic and is susceptible to external factors. However, when performed under the same temperature difference ($\Delta T = 30 - 25^\circ\text{C}$), the supersaturation ratios of the compounds are different as a result of the different gradients in their solubility temperature profile. Based on our theoretical derivations of induction time and supersaturation ratio, higher supersaturation will lead to lower induction time. This trend was seen on alanine and Cbz-alanine, whereas Boc-alanine did not nucleate at all. Any conclusions made on the effects of the protecting group on induction time will be unsatisfactory, as probing for more repeats and a controlled experimental setup. Otherwise, we can only conclude that Cbz-alanine is more likely to induce quicker than alanine due to its consistently smaller induction time. To understand this effect, we looked at their pre-exponential factor, A and thermodynamic factor, B . It was found that the presence of benzyloxycarbonyl increases A and B of alanine. The benzyloxycarbonyl group increases the surface area of the molecule, increasing the likelihood of a successful attachment between a building unit and the nucleus. This is counteracted by the increase in energy barrier that the bulky protecting group introduces, reducing

the likelihood that a stable solid-liquid interface can be formed. The result of this experiment shows large deviations, which only allows for a quantitative analysis of the result. For a qualitative analysis, this experiment should be conducted with more repeats to account for the stochastic nature of induction time and other external factors that can affect the deviations.

At 25°C, alanine solubility was not affected by the presence of Boc-alanine and Cbz-alanine impurity. This is also the same when alanine impurity was added into a solution of Boc-alanine and Cbz-alanine, where no statistically significant effects can be observed. There are positives and negatives to this, as it is more difficult to detect the presence of an impurity based on its solubility data alone. Regardless, the solubility performance of alanine would remain unchanged, and no alterations to the design process have to be made.

Surprisingly, it was found that the solubility of Cbz-alanine in a solution of DI water + alanine is 2-5 times higher than in a solution of pure DI water. This result was seen under different alanine concentrations in DI water and under different temperatures, signifying that alanine promotes the dissolution of Cbz-alanine. A solution of DI water and alanine changes the overall polarity of water, allowing for better hydrogen bond interactions with Cbz-alanine, increasing the apparent solubility of Cbz-alanine. This is a significant problem when synthesising alanine, as Cbz-alanine may dissolve in the solution of alanine + DI water, increasing impurity presence and affecting the total yield, similar to the case mentioned in [24].

We also explored the external effects on solubility and induction time to justify our methodologies. It was found that recycling samples maintains the stochasticity of induction time. This means induced solution samples can be reheated and re-cooled to collect another set of data for the same compound instead of producing a completely new sample. We also compared how agitation may affect the results of four data, and it was shown that better agitation leads to better hydrodynamics, resulting in better mass transfer. This does not mean that higher RPMs lead to better results, as we have seen that an optimum operating window can be achieved. This was not further explored in this paper and is a potential outlook for further research to aid in better data accuracy for the solubility of components.

Outlook. This paper has been an attempt to investigate critical compounds involved in the synthesis of a long peptide chain of alanine. However, there are potential future outlooks stemming from this experiment to further contribute to this search. More number of repeats for the induction time of alanine, Boc-alanine and Cbz-alanine is to be done to properly obtain the qualitative values of the pre-exponential factor, A and the thermodynamic factor, B . Once a consistent value is obtained, the value of γ_{int} can be calculated and compared with the value obtained in this paper. The induction time experiment can also be done at differing temperature ranges but at the same supersaturation ratios to compare the temperature driving force, ΔT , required to achieve a desired induction time. For impurities, there is an option to explore how the impurities affect the induction time of alanine, as this was not explored in this paper. Hence, the effects of the impurities on the pre-exponential and thermodynamic factors. It is also worth noting that studying the crystal habits in the presence of impurities should also be of interest to ensure any forms of polymorphism are generated in the mixture.

Table 4. Table of Symbols and Units

Symbol	Description	Unit
A	Pre-exponential Factor	s
B	Thermodynamical Factor	-
c	Concentration	mol/mL
c^*	Saturated Concentration	mol/mL
D	Diameter	m^2
f^*	Attachment Frequency	s^{-1}
J	Nucleation Rate	s^{-1}
K	Boltzmann Constant	J/K
k_m	Mass Transfer Constant	m/s
m_A	Mass of Solute	mg
m_I	Mass of Impurity	mg
m_s	Mass of Solution	mg
m_T	Total Mass	mg
m_{solv}	Mass of Solvent	mg
M	Molecular Weight	g/mol
n_i	Number of Ions/Species Formed in Solution	-
p	Pressure	Pa
P_t	Probability of Induction Time to Occur	-
R	Gas Constant	J/mol.K
R_0	Induction Time at 0 th Recycle	min
R_n	Induction Time at n th Recycle	min
r_c	Critical Radius	m
Re	Reynolds Number	-
ρ_s	Density of Solids	kg/m ³
S	Supersaturation Ratio	-
Sc	Schmidt Number	-
t_i	Induction Time	s
T	Temperature	K
U	Fluid Velocity	m/s
V	Volume	m ³
z	Zeldovich Factor	-
ΔG_c	Critical Gibbs Free Energy	J
ΔG_r	Gibbs Free Energy for Surface Term	J
ΔG_V	Gibbs Free Energy for Volume Term	J
γ_{int}	Interfacial Surface Energy	J/m ²

ACKNOWLEDGMENTS. This paper was done with the guidance of Jianing Li, from Imperial College London. Heng Group from Imperial College London has also given advice and guidance during the writing of this paper, notably Deniz Etit and Dr. Isha Bade

References

- [1] Amblard, M., Fehrentz, J.-A., Martinez, J., and Subra, G. (2007). Methods and protocols of modern solid-phase peptide synthesis. *ChemInform*, 38(30).
- [2] Brandel, C. and ter Horst, J. (2015). Measuring induction times and crystal nucleation rates. *Faraday Discussions*, 179:199–214.
- [3] Cardoso, M., Ribeiro, S., Nolasco, D., et al. (2016). A polyaniline peptide derived from polar fish with anti-infectious activities. *Scientific Reports*, 6:21385.
- [4] Dąbrowska, N. (2017). *Influence of Impurities on the Crystallization of Pharmaceutical Materials*. PhD thesis. Thesis. p.4.
- [5] Ely, D. R., García, R. E., and Thommes, M. (2014). Ostwald–freundlich diffusion-limited dissolution kinetics of nanoparticles. *Powder Technology*, 257:120–123.
- [6] International Conference on Harmonization (1996). Guideline for industry structure and content of clinical study reports. [online] Available at: <https://www.fda.gov/media/71271/download>.
- [7] IUCr Commission (n.d.). Crystal. Accessed: 2025-12-01.
- [8] Izmailov, A. and Myerson, A. (2000). Diffusion in supersaturated solutions: Application to the case of supersaturated protein solutions. *The Journal of Chemical Physics*, 112(9):4357–4364.
- [9] Keshavarz, L., Steendam, R. R. E., Blijlevens, M. A. R., Pishnamazi, M., and Frawley, P. J. (2019). Influence of impurities on the solubility, nucleation, crystallization, and compressibility of paracetamol. *Crystal Growth & Design*, 19(7):4193–4201.
- [10] Klug, D. L. (1993). Handbook of industrial crystallization. In Myerson, A. S., editor, *Handbook of Industrial Crystallization*, page 65. Butterworth-Heinemann, Boston.
- [11] Kubota, N. (2001). Effect of impurities on the growth kinetics of crystals. *Crystal Research and Technology*, 36(8/10):749–769.
- [12] Lian, Z., Wang, N., Tian, Y., and Huang, L. (2021). Characterization of synthetic peptide therapeutics using liquid chromatography–mass spectrometry: Challenges, solutions, pitfalls, and future perspectives. *Journal of the American Society for Mass Spectrometry*, 32(8):1852–1860.
- [13] Lu, J. X., Tupper, C., Gutierrez, A. V., et al. (2025). *Biochemistry, Dissolution and Solubility*. StatPearls Publishing, Treasure Island (FL). Updated 2022 Sep 12. Available from: StatPearls [Internet]. Accessed: 2025-12-08.
- [14] Mullin, J. W. (2013). *Crystallization*. Butterworth-Heinemann Limited, 5 edition. 20–35 pages.
- [15] National Center for Biotechnology Information (2025a). Pubchem compound summary for cid 5950, l-alanine. Accessed: 2025-12-01.
- [16] National Center for Biotechnology Information (2025b). Pubchem compound summary for cid 637606, boc-d-alanine. Accessed: 2025-12-01.
- [17] National Center for Biotechnology Information (2025c). Pubchem compound summary for cid 736104, n-carbobenzyloxalanine. Accessed: 2025-12-01.
- [18] Ng, K., Gani, R., and Dam-Johansen, K. (2006). *Chemical Product Design: Towards a Perspective through Case Studies*. Elsevier.
- [19] Nicholson, S., Roberts, K., Izumi, T., and Lai, X. (2023). Characterization of mass transfer within the crystal-solution boundary layer of l-alanine {120} faces using laser interferometry during growth and dissolution. *Crystal Growth & Design*, 23(4):2755–2769.
- [20] Perez-Sala, D., Parrilla, R., and Ayuso, M. S. (1987). Key role of l-alanine in the control of hepatic protein synthesis. *Biochemical Journal*, 241:491–498. Printed in Great Britain.
- [21] Sharma, A., Mehta, V., Rani, S., et al. (2022). Biomedical applications of l-alanine produced by *Pediococcus acidilactici* bd16 (alad+). *Applied Microbiology and Biotechnology*, 106:1435–1446.
- [22] Shiau, L. D. (2022). Comparison of the nucleation kinetics obtained from the cumulative distributions of the metastable zone width and induction time data. *Molecules*, 27(9):3007.
- [23] Tymecka, D. and Misicka, A. (2025). Solution phase peptide synthesis: The case of biphalin. In *Methods in molecular biology*, pages 143–154.
- [24] Urwin, S. J., Yerdelen, S., Houson, I., and ter Horst, J. H. (2021). Impact of impurities on crystallization and product quality: A case study with paracetamol. *Crystals*, 11(11):1344.
- [25] Yang, H. and Rasmuson, (2013). Nucleation of butyl paraben in different solvents. *Crystal Growth & Design*, 13(10):4226–4238.

Evaluating Actor-Critic Reinforcement Learning Algorithms for Process Control Applications in a CSTR

Sanjiv Jadav and Yuta Yoshizaki

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Proportional-integral-derivative (PID) control remains the most widely used control strategy within the chemical industry due to its reliability and proven effectiveness in minimising the error between the process variable and its set-point. However, achieving effective PID control performance often requires extensive manual tuning, and even well-tuned PID controllers can struggle to maintain satisfactory performance in highly nonlinear systems. Model predictive control (MPC) offers an established industrial solution to address system nonlinearity, though its effectiveness relies heavily on the availability of an accurate and representative process model. Therefore, this research investigates the use of reinforcement learning (RL) as an advanced process control solution. Unlike PID control and MPC, RL does not rely on fixed tuning parameters or explicit process models, and so allows the controller to learn optimal control behaviours directly from data. The first component of this work benchmarks an Advantage Actor-Critic (A2C) algorithm-trained controller against a conventionally tuned PID controller for a simple flow control application. The second component of this work investigates the ability of a Deep Deterministic Policy Gradient (DDPG) algorithm-trained controller to generate favourable control actions to achieve stable operation of a nonlinear and multivariable continuous stirred tank reactor (CSTR). The results demonstrate the potential of RL as an advanced process control strategy, yielding promising outcomes across both components.

1. Introduction

Process control is a critical aspect of operating plants within the chemical manufacturing and broader industrial sectors, as it is key in ensuring processes are run safely, reliably and consistently^[1]. In the chemical manufacturing industry, traditional control strategies are typically based on classical closed-loop feedback techniques, most notably PID control, due to its ease of implementation, disturbance-rejection and set-point tracking capabilities^[2]. However, for a given set of PID tuning parameters, more formally known as the PID gains, their effectiveness is generally limited to a narrow operating range where the process dynamics remain approximately linear^[3]. This limitation has prompted the use of gain scheduling, which involves designing multiple PID controllers for different operating regions and switching between them according to the current process conditions^[4]. Despite this, PID control still presents challenges when applied to highly nonlinear and multivariable processes^[2]. This is because its control actions are solely driven by the instantaneous error, and it lacks the capability to comprehend the underlying process dynamics or the impact of interactions among different control loops^[3].

These challenges have promoted the exploration of more advanced control techniques, namely Model Predictive Control (MPC), which has been successfully implemented in industry^[5]. More recently, reinforcement learning (RL) has emerged as a promising model-free solution for advanced process control^[6], as it offers the ability to learn effective control policies directly from data without requiring an explicit model of the system dynamics.

This gives RL a distinct advantage over MPC in systems, where dynamics are difficult to model accurately^[6].

The recent research highlights the use of RL for process control, with studies highlighting the successful integration of RL with PID control^[8], thereby showing that while PID remains well-established in industry, its performance can be substantially enhanced through data-driven optimisation. However, many of the RL methods reported in literature rely on online interactions^[9] with the process, which can be impractical for most industrial settings. Pure online RL requires exploratory actions that may drive the process outside an acceptable operating range^[10], posing risks to both safety and production. Therefore, an offline RL approach, whereby the agent trains on historical process data, offers a safer and more cost-effective way to develop improved control strategies for industrial applications^[10].

This work investigates RL as an advanced process control strategy for a CSTR system. The first component of this work benchmarks an A2C-based controller against a conventional Lambda-tuned PID controller for the control of the inlet flow-rate to a CSTR. The second component of this work examines whether a DDPG model can directly manipulate the operating conditions of a nonlinear CSTR by training on a first principles reactor model with bounded actuator limits.

2. Background

2.1. Reinforcement Learning

Reinforcement learning is a branch of machine learning in which an agent learns to perform actions through interactions with an environment^[11]. At each discrete time step t , the agent observes the current system state s_t and selects an action A_t according to its policy $\pi(s_t/A_t)$ and a reward R_t feedback from the environment $\mathbb{E}$, with the objective of learning a policy π^* that maximises the expected cumulative return^[11].

$$J(\pi) = E_{\pi} \left[\sum_{t=0}^T \gamma^t R_{t+1} \right] \quad (1)$$

$$\pi^* = \arg \max_{\pi} J(\pi) \quad (2)$$

, where $J(\pi)$ represents the performance measure of the policy π , γ is the discount factor applied to future rewards, and π^* is defined as the optimal policy that maximises $J(\pi)$.

The interaction between the agent and the environment can be described using a Markov Decision Process (MDP), defined by the tuple $\langle \mathcal{S}, \mathcal{A}, \mathbb{P}, \mathbb{R}, \gamma \rangle$, where $\mathcal{S}$ denotes the set of possible states, $\mathcal{A}$ denotes the set of admissible actions, $\mathbb{P}(s_{t+1}|s_t, A_t)$ describes the system transition dynamics, $\mathbb{R}(s_t, A_t)$ specifies the reward function and γ is the discount factor applied to the reward as mentioned previously. The MDP provides a mathematical framework within which the policy operates, capturing how actions influence states and rewards at each time step^[11].

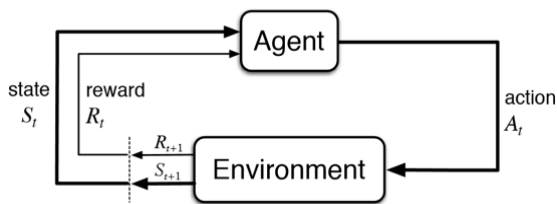


Figure 1: Agent-Environment interaction in a Markov Decision Process (reproduced from Sutton & Barto ^[11])

Beyond MDP, reinforcement learning methods can be categorised by value-based, policy-based or both. Value-based methods learn an estimate of the expected return for each state-action pair and derive the policy from this. In contrast, policy-based methods directly learn a parametrised policy that maps states to actions^[11]. Actor-critic algorithms, such as A2C and DDPG used in this work, combine both approaches by learning a policy (actor) and a value function (critic) simultaneously^[11]. This hybrid structure allows the agent to benefit from more stable

and efficient learning, making it a practical choice for the nonlinear control problems considered in this research.

2.2. PID Control

The PID controller remains the foundation of nearly all basic control applications within the chemical industry^[1]. It functions by calculating an error value defined as the difference between a target set-point and the measured process variable, and then applies corrective actions to the error through proportional, integral and derivative terms^[1]. The proportional action provides immediate correction to deviations from the set-point by producing a controller output directly proportional to the instantaneous error, scaled by the proportional gain K_p ^[2]. The integral action removes any residual steady-state errors by producing a controller output based on the accumulation of past errors, which is scaled by the integral time constant τ_i ^[2]. The derivative action helps dampen any fluctuations and improve stability by producing a controller output proportional to the rate of change of the error, which is scaled by the derivative time constant τ_d ^[2]. These corrective actions are computed at discrete sampling intervals with the controller updating its output at each sample using the latest error measurement ^[1]. The PID controller defined through a discrete sampling time interval Δt is given by Equation 3.

$$u_t = K_p e_t + \frac{K_p}{\tau_i} \sum_{t=0}^T e_t \Delta t + K_p \tau_d \frac{e_t - e_{t-1}}{\Delta t} \quad (3)$$

, where u_t is the controller output, $e_t = x_t^{sp} - x_t$ is the error, x_t^{sp} is the set-point for the process variable x and Δt is the sampling period.

However, in order to achieve satisfactory control performance from a PID controller, the gain values K_p , τ_i and τ_d , must be adequately tuned. Poor tuning can result in sluggish response, excessive oscillations and process instability^[2]. Many of the well-established tuning methods, such as the Ziegler-Nichols method and Lambda tuning, rely on online experimentation, which can be unsafe, especially for processes with dynamics that cannot be modelled so easily^[2]. Consequently, a reinforcement learning approach training on historical process data could provide a safer and less time-consuming alternative to PID control^[9].

3. Methodology

3.1. Case Study Overview

This research is structured into two complementary case studies that together evaluate the potential of RL

as an advanced control strategy for CSTR systems. The first case study investigates an A2C-based controller for regulating the inlet flow-rate to a CSTR and benchmarking its performance with a conventional PID controller. Flow control provides an ideal test environment for this benchmarking exercise, as it represents a task that PID control is already highly effective in, while also evaluating whether an RL-based controller's explicit handling of constraints enhances or compromises set-point tracking performance and stable convergence. A static transfer function, derived from empirically obtained step-response data, is used as the environment for the A2C agent to train within. The PID controller performance is evaluated directly on the physical CSTR rig, where flow regulation is achieved by a diaphragm pump. This allows the evaluation of whether a model-free actor-critic agent can learn the fundamental control behaviours shown by the industry-established PID controller.

The second case study explores the use of RL to manipulate the operating conditions of a nonlinear CSTR through a model-based training approach. A DDPG agent is employed to assess whether the flow control capability demonstrated in the first case study can be extended to a highly nonlinear and multivariable CSTR. The DDPG agent is trained on a first principles reactor model to generate continuous control actions subject to bounded actuator limits, enabling it to learn how changes to the inlet conditions influence the reactor dynamics. Therefore, this case study evaluates the ability of a DDPG agent to learn stable control behaviours within a highly nonlinear and multivariable system.

Therefore, the two case studies evaluated in this work provides a complementary evaluation of RL within chemical process control, with the first case study examining whether an RL agent can learn the fundamental control behaviours associated with flow-control relative to a PID controller. The second case study extends the learning of an RL agent to manipulating the operating conditions of a nonlinear and multivariable CSTR using a model-based training approach.

3.2. Case Study 1: A2C-based Flow Controller

The first case study benchmarks the performance of an A2C-trained controller against a conventional PID controller, for the inlet flow to a 2L CSTR. The PID controller was tuned using the Lambda-tuning method on pre-collected step-response data, where the dynamics were approximated as first-order plus deadtime (FOPDT) and characterised by the process gain, time constant and time delay^[2]. The PID gains (K_p , τ_i , τ_d) obtained from the Lambda-tuning method are then experimentally computed on the physical CSTR rig.

The same FOPDT transfer function makes up the training environment for the A2C agent. Within this environment, the agent follows an actor-critic architecture, where the actor network outputs continuous control actions and the critic network estimates the value function $V(s)$ of the control actions, and subsequent policy updates are guided by an advantage term^[11] given by Equation 4.

$$a_t = r_t + \gamma V(s_{t+1}) - V(s) \quad (4)$$

The advantage term quantifies whether an action performed better or worse than expected, which allows the agent to iteratively improve its policy using gradient-based updates.

The advantage term also incorporates the reward at each time step and in this work the reward was deliberately designed to provide strong corrective action during learning. By imposing strict penalties on the tracking error and excessive control effort, the reward function encourages policy refinement at a much faster rate, which mitigated the effect of limited training time. To further support stable learning, the agent was trained using a low learning rate, 2×10^{-4} , which helps prevent large oscillations in the policy during training and ensures that improvements to the actor and critic networks are made more gradually.

$$\begin{aligned} r_t = & \frac{e_t}{x^{sp}} - 0.0001|A_{t+1} - A_t| - 0.0005 \left| \frac{e_{t+1} - 2e_t + 2e_{t-1}}{(\Delta t)^2} \right| \\ & + \begin{cases} 20, & e_t < 0.01 \\ 10, & e_t < 0.05 + 2 \text{ (if } e_t < 0.02 \text{ \& } t > 50) \\ 5, & e_t < 0.1 \end{cases} \\ & + 3, \text{ if } e_t < 0.05 \text{ \& } \left| \frac{e_{t+1} - e_t}{\Delta t} \right| \text{ \& } \left| \frac{e_{t+1} - 2e_t + 2e_{t-1}}{(\Delta t)^2} \right| < 0.5 \\ & - \min \left(0.1 \times \frac{t-100}{100}, 5 \right) \text{ (if } t > 100 \text{ \& } |e_{t+1}|) \end{aligned}$$

Figure 2: Reward function used to train the A2C Controller

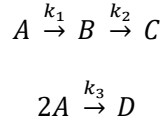
The reward function, as shown in Figure 2, was designed to guide the agent's behaviour in a way that mirrors classical control objectives, such as set-point tracking, stable convergence, minimised control effort and fast response times. The proportional-like term, given by $\frac{e_t}{x^{sp}}$, directly penalises the instantaneous tracking error and encourages the agent to minimise deviations from the set-point. To prevent aggressive changes in control actions, the term $-0.0001|A_{t+1} - A_t|$, penalises large actuator movements and promotes smoother control responses. The reward function also discourages slow settling times through a time-dependent penalty, as given by $-\min \left(0.1 \times \frac{t-100}{100}, 5 \right)$, which becomes increasingly negative when the system has not settled after a specified horizon. Set-point overshoot is mitigated by a piecewise discrete reward formulation in which the reward increases as the process variable

approaches the set-point, thereby disincentivising overshoot. Additionally, the inclusion of first- and second-order derivative terms promote stable and well-damped closed-loop behaviour by penalising rapid changes and curvature in the tracking error. To ensure these terms interacted sensibly with the main tracking objective, the reward parameters were chosen such that each term within the reward function had a clear relative weighting and remained on a similar scale to the base error term.

3.3. Case Study 2: DDPG Model for CSTR Control

The second case study assesses whether the flow control capability demonstrated by the A2C-based controller can be extended by a DDPG agent to the control of a highly nonlinear and multivariable CSTR system. The environment used in this case study is based on an open-source CSTR control code^[12], where the CSTR reaction exhibits van de Vusse kinetics^[7]. The reactor dynamics are described by a system of coupled first-order ordinary differential equations (ODEs), resulting from the mass- and energy-balances of the reactor and cooling jacket.

The van de Vusse kinetics are used to describe the following reaction network:



, where the concentrations of species A and B are tracked.

The following system of ODEs form the training environment for the DDPG agent.

$$\begin{aligned} \frac{dC_A}{dt} &= \frac{F[C_{A0} - C_A]}{V} - k_1(T_R)C_A - k_3(T_R)C_A^2 \\ \frac{dC_B}{dt} &= \frac{F[C_{B0} - C_B]}{V} + k_1(T_R)C_A - k_2(T_R)C_B \\ \frac{dT_R}{dt} &= \frac{F[T_{in} - T_R]}{V} + \frac{K_w A_R [T_w - T_R]}{\rho C_p V} \\ &\quad + \frac{1}{\rho C_p} [k_1(T_R)C_A \Delta H_1 + k_2(T_R)C_B \Delta H_2 \\ &\quad \quad \quad + k_3(T_R)C_A^2 \Delta H_3] \\ \frac{dT_k}{dt} &= \frac{\dot{Q} + K_w A_R [T_k - T_R]}{(MC_p)_k} \end{aligned} \quad (5)$$

, where the rate constant for reaction i is k_i and is given by Equation 6.

$$k_i(T) = k_{0,i} e^{-E_{Ai}/RT_R} \quad (6)$$

The fixed parameters within the system of ODEs are located in Table 1 and 2.

Reaction	$k_{0,i}$	E_{Ai}/R	ΔH_i
$A \rightarrow B$	$1.287 \times 10^{12} \text{ h}^{-1}$	-9758.3 K	+4.2 kJ mol ⁻¹
$B \rightarrow C$	$1.287 \times 10^{12} \text{ h}^{-1}$	-9758.3 K	-11 kJ mol ⁻¹
$2A \rightarrow D$	$9.043 \times 10^{12} \text{ L mol}^{-1} \text{ h}^{-1}$	-8560 K	-41.85 kJ mol ⁻¹

Table 1: Reaction kinetic parameters

Parameters	Symbols	Values	Units
Inlet Flow-rate	F	-	L h ⁻¹
Reaction mixture density	ρ	0.9342	Kg L ⁻¹
Reaction mixture heat capacity	C_p	3.01	kJ kg ⁻¹ K ⁻¹
Thermal conductivity of the reactor jacket wall	K_w	4032	kJ m ⁻¹ h ⁻¹ K ⁻¹
Reactor heat exchange area	A_R	0.215	m ²
Reactor volume	V_R	10	L
Heat capacity of the reactor jacket	$(MC_p)_w$	10	kJ K ⁻¹

Table 2: Physiochemical properties and reactor dimensions

The underlying assumptions associated with the coupled mass- and energy-balances are listed as follows:

- Constant reaction mixture density
- Constant reaction mixture heat capacity
- Constant reactor level

These assumptions simplify the governing mass- and energy-balances, while still preserving the characteristic dynamic behaviours of a CSTR.

In this study, the inlet flow-rate and jacket coolant temperature serve as the manipulated variables used to control the concentrations of species A and B and the reactor temperature. Both manipulated variables influence all controlled states through the coupled-mass- and energy-balances, which results in an integrated multivariable system. These multivariable interactions present challenges for conventional PID control, as these independent single-input-single-output (SISO) loops cannot compensate for coupling between control variables.

The training setup uses a DDPG agent to control this nonlinear and multivariable CSTR model. Within this environment, the DDPG agent follows an actor-critic architecture, where the interaction between the actor and critic networks are driven by the action's value

function $Q(s_t, A_t)$, which estimates the expected return for taking an action A_t in state s_t ^[11]. The critic is trained using a temporal difference target which acts as the learning signal for evaluating the quality of the agent's actions^[11]. The temporal difference target is defined in Equation 7.

$$y_t = r_t + \gamma Q_{target}(s_{t+1}, \pi^*(s_{t+1})) \quad (7)$$

The resulting error given as ' $y_t - Q(s_t, A_t)$ ', is the analogous to the advantage term in the A2C algorithm formulation.

Both the actor and critic networks are trained using an Adam Optimiser with a fixed learning rate of 3×10^{-4} . A fixed learning rate is chosen to balance the stability and speed of convergence, as the DDPG algorithm is sensitive to aggressive parameter updates.

At each time step, the environment provides the reaction states along with their deviations from the set-point. The control policy outputs two continuous control actions, corresponding to the inlet flow rate and the inlet coolant temperature, which are rescaled into physical actuator inputs before being applied to the CSTR simulation. The simulator then computes the next state using the differential equations outlined in Equation 6. This closed-loop interaction forms the basis of the DDPG agent's learning process, as each transition allows the agent to update its understanding of how the manipulated variables influence the reactor dynamics.

The reward function, as shown in Figure 3, is defined as the negative summation of the squared tracking errors of each control variable. This formulation of the reward encourages the agent to minimise deviations across all states simultaneously and therefore learn coordinated multivariable control behaviour. This reward is accumulated over the course of an episode, after which the episode terminates according to predefined conditions.

$$r_t = -[(e_t^{c_A})^2 + (e_t^{c_B})^2 + (e_t^{T_R})^2 + (e_t^{T_k})^2]$$

Figure 3: Reward function used to train the DDPG Controller

The episodes terminate either when the states remain sufficiently close to the set-point for several consecutive steps, as indicated by '*Done: True*' or when the maximum episode length of 3000 is reached.

4. Results and Discussion

4.1. Case Study 1: A2C-based Flow Controller

The performance of the A2C-trained controller was evaluated against a Lambda-tuned PID controller on a 2L CSTR system. Both controllers were tested across a range of flow-rate set-points (between 0.2 – 0.8 L/min) to assess their closed-loop behaviour. To quantitatively compare the performances of each controller, the following metrics were used: steady-state error, average and maximum absolute error, integral of absolute error (IAE), integral of time-averaged error (ITAE) and settling time. These metrics capture the accuracy of the controllers as well as the characteristics of their responses to transient behaviours.

4.1.1. PID Controller

To determine whether the A2C agent could learn the fundamental control behaviours associated with flow control, its performance was benchmarked against a Lambda-tuned PID controller. Table 3 shows the performance of the PID controller when tested on the physical CSTR rig.

	Lambda-tuned PID Controller ($K_p = 1.4406 \text{ L/min}$, $\tau_i = 6.798 \text{ min}$, $\tau_d = 0.000 \text{ min}$)				
Flowrate Setpoint (L/min)	0.2	0.3	0.4	0.6	0.8
Final Flow (L/min)	0.2009	0.3004	0.4005	0.6033	0.7994
Steady-state Error (L/min)	0.0009	0.0004	0.0005	0.0033	-0.0006
Average Absolute Error (L/min)	0.0043	0.0139	0.0205	0.0185	0.0962
Maximum Absolute Error (L/min)	0.1120	0.2350	0.3260	0.4270	0.7880
IAE	1.4980	3.4326	5.2257	10.6833	16.3155
ITAE	170.9808	137.6593	164.1479	250.3296	334.487
Settling Time (s)	263.0	41.1	46.9	49.9	62.9

Table 3: Results summary table for PID controller

The data shown in Table 3 highlights that the Lambda-tuned PID controller was capable of achieving successful set-point tracking across the full range of tested set-points. This outcome is to be expected because this type of simple flow process is one for which PID control has demonstrated in industry to provide robust and reliable performance. The noticeably poorer control performance observed for the 0.2 L/min set-point was to be expected due to the known limitations of diaphragm pumps operating at lower flow-rates. However, the inclusion of this flow set-point in this study was intended to validate whether the A2C agent had learned a control policy compatible with the real pump behaviour, given that it was trained offline and could potentially adopt unrealistic behaviours.

4.1.2. A2C-trained Controller

The performance of the A2C-trained controller is summarised in Table 4.

A2C Controller					
Flowrate Setpoint (L/min)	0.2	0.3	0.4	0.6	0.8
Final Flow (L/min)	0	0.0396	0.0821	0.5982	0.7998
Steady-state Error (L/min)	0.2	0.2604	0.335	0.0019	0.0002
Average Absolute Error (L/min)	0.1938	0.2549	0.3277	0.0345	0.017
Maximum Absolute Error (L/min)	0.2	0.3485	0.4	0.6	0.8
IAE	58.13	76.43	98.27	10.33	5.06
ITAE	8977.54	11704.45	15058.29	641.46	196.92
Undershoot Percentage (%)	99.5	97.8	98.9	99	99.2
Settling time (s)	-	-	-	170.5	80.8

Table 4: Results summary table for A2C-trained controller

The data shown in Table 4 highlights the control performance achieved by the A2C-trained controller. For the 0.6 and 0.8 L/min flow-rates, the A2C-trained controller successfully converged to the set-point values, therefore demonstrating that the agent had learned a control policy capable of tracking the error. However, the settling times associated with these convergences are 80.8 and 170.5 seconds, which in comparison to the 49.9 and 62.9 seconds achieved by the PID controller, show much slower convergences despite an explicit penalty for slow response in the reward function. This indicates the need for further refinement of the penalty weightings to better balance the trade-off between minimising control effort and achieving faster convergence.

For the intermediate set-points, 0.3 and 0.4 L/min, the controller failed to converge and resulted in large steady-state errors. This behaviour can be attributed to the penalty weightings in the reward function as for these set-point values the initial error is too low such that the reward obtained from further reducing it is outweighed by the penalty associated with excessive control effort. As a result, the agent learns to accept a large steady-state error rather than incurring an additional penalty for control effort.

For the 0.2 L/min set-point, the A2C-trained controller produced a final flow of zero, indicating an outright failure to regulate the process at this flow-rate. This outcome offers confidence that the agent did not adopt unrealistic behaviours, and instead learned a control policy broadly compatible with the real pump's low-flow limitations, despite being trained entirely offline using a static transfer function.

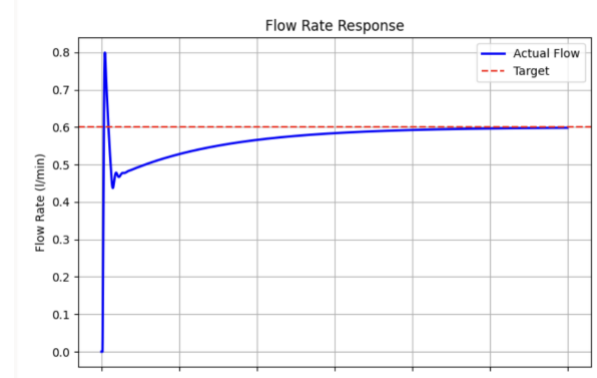


Figure 4: A2C Controller Flow Response for 0.6 L/min set-point

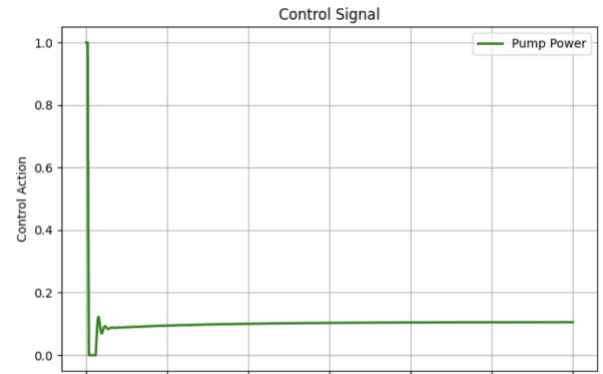


Figure 5: A2C Controller Control Signal response for 0.6 L/min set-point

Figure 4 and 5 shows the closed-loop response of the A2C-trained controller for the 0.6 L/min set-point. The flow trajectory initially overshoots the target before undershooting, after which it gradually converges toward the set-point. Between the undershoot and the gradual convergence, small-amplitude oscillations can be observed however as the pump power smoothens, they are dampened. Overall, this profile is consistent with the description of the reward function for the higher set-points, whereby the reward for error reduction dominates the penalty for excessive control effort and thus enabling convergence.

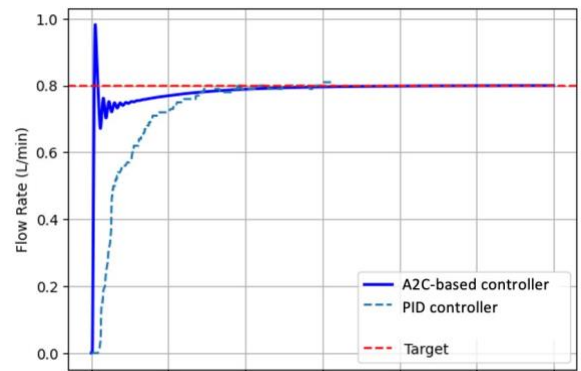


Figure 6: Comparison of flow-rate tracking performance between the A2C-trained controller and a PID controller at 0.8 L/min set-point

Figure 4 presents the step-response comparison between the A2C-trained controller and a Lambda-tuned PID controller for regulating the inlet flow-rate to a 2L CSTR vessel. The two controllers exhibit distinctly different transient behaviours, although they both eventually converge to the set-point.

For the A2C-trained controller, the flow-rate response exhibited a sharp overshoot followed by small decaying oscillations before eventually converging toward the set-point. In contrast, the Lambda-tuned PID controller response tracked smoothly throughout with negligible overshoot or oscillations. This reflects that the A2C agent had not fully learned the damping characteristics needed to smoothen transient behaviours. This is likely due to conservatism in the agent's learning, which causes it to avoid making the actuator movements needed to dampen the transients, due to the associated penalty.

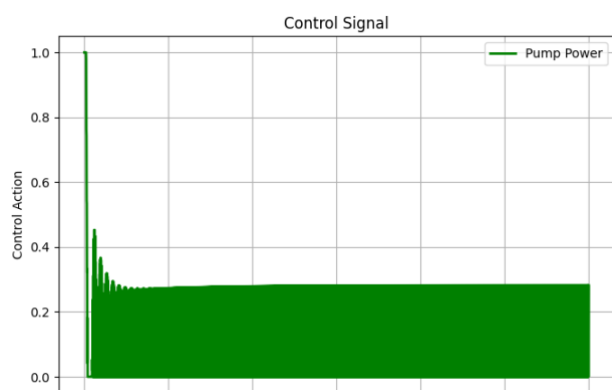


Figure 7: A2C-trained controller signal response for 0.8 L/min set-point

Figure 7 shows the control signal response from the A2C-trained controller for the 0.8 L/min set-point. The response observed is markedly different to that observed for the 0.6 L/min set-point or from the corresponding PID controller output.

The learned control behaviour to achieve the 0.8 L/min set-point exhibits an ‘on-off’-like control, with the pump motor rapidly switching between zero and non-zero power values. This behaviour is indicative of a more aggressive approach to set-point tracking and stable convergence, which the agent likely adopted in response to the large initial error and in the absence of any explicit penalty on high-frequency actuation.

For the physical process, this rapid ‘on-off’ actuation is generally undesirable, even for a diaphragm pump, which inherently generates flow through diaphragm pulsations, as the switching frequency exceeds the pump’s maximum operating frequency.

In summary, the A2C-trained controller performances were encouraging as the agent was

capable of demonstrating set-point tracking, albeit not for the full range of tested set-points. This indicates the need for further refinement of the reward function, particularly the penalty weightings, to better balance the trade-off between set-point tracking, control effort and settling times. Improved penalty weightings would make it plausible for the A2C-trained controller to generalise its tracking behaviour across the full range of set-points, even within the offline environment. Overall, these results demonstrate the potential of RL for process control while also highlighting that strong performance requires careful reward and penalty design.

4.2. Case Study 2: DDPG Model for CSTR Control

The performance of the DDPG-trained controller was evaluated on a nonlinear and multivariable CSTR environment to assess whether the agent could learn a stable control policy capable of set-point tracking multiple control variables. As outlined in Section 3.3., the agent interacts with the reactor model through continuous actuator inputs, and its behaviour develops over successive training episodes. To characterise this learning progression, the results are presented in terms of the concentration trajectories of species A and B, and the temperature trajectories of the reactor and cooling fluid, across the full training horizon. The trajectories are coupled with the corresponding reward profiles to illustrate how improvements in tracking performance translate to higher returns. Collectively, the results provide a clear indication of the improvement in the learned policy and the associated improvement in the controller’s closed-loop behaviour within a nonlinear environment.

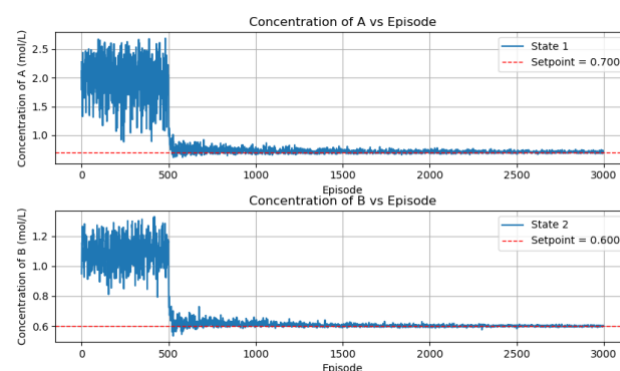


Figure 8: Concentration trajectories for species A and B

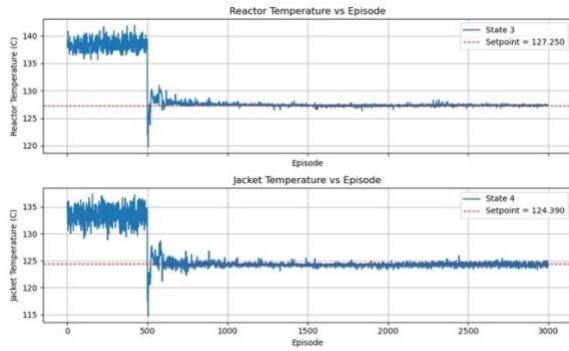


Figure 9: Temperature trajectories for the reactor and the cooling fluid

The state trajectories shown in Figures 8 and 9, highlight a clear learning progression in the DDPG agent. Each point in the trajectory represents the steady-state value of the process variable achieved at the end of an episode following a step-response experiment. Across all four reactor states, the initial episodes are characterised by large fluctuations and significant deviation from the set-point, reflecting the initial phase of the agent performing random control actions to explore the inherent nonlinearity associated with the van de Vusse reaction system. The training phase starts at the 500th episode where a distinct transition can be observed, after which the trajectories converge to within a narrow-band of the states' respective set-points and exhibit significantly less fluctuations. The simultaneous stabilisation of the concentration and temperature states at the 500 episode-mark, indicates that the agent has started to learn an effective state-action mapping consistent with the reactor's mass and energy-balances. This demonstrates the capability of a DDPG-trained controller to provide effective closed-loop control for both a nonlinear and multivariable process.

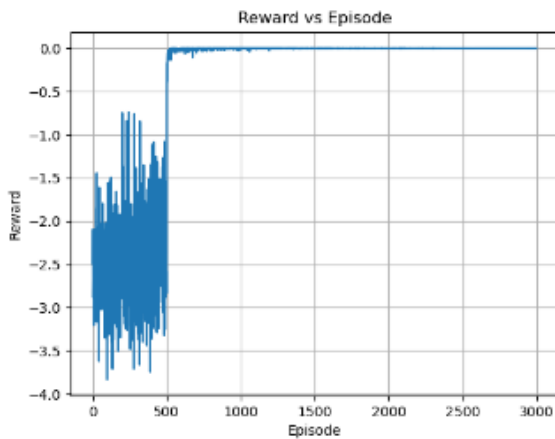


Figure 10: Reward profile over the full training horizon

The reward profile over the full training horizon, shown in Figure 10, mirrors the state behaviours and reflects the overall progression of the learning process. During the early episodes, the reward

fluctuates around large negative values due to inadequate set-point tracking and random policy initialisation. A sharp transition occurs at 500th episode, after which the reward rises rapidly towards zero and remains stable around this point. This sharp transition and stabilisation around a reward value of zero, is indicative of the agent learning a policy that achieves adequate set-point tracking across all reactor states.

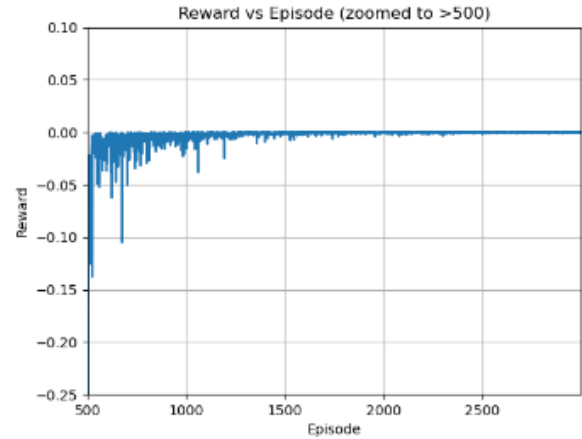


Figure 11: Zoomed-in reward profile for episodes > 500

The zoomed-in reward profile, shown in Figure 11, provides additional insight into the refinement of the control policy after the initial convergence. It can be seen that directly after the transition at when the training phase begins, there are still subtle fluctuations, which indicate occasional deviations arising from ongoing exploration and policy updates. However, as the number of episodes increases beyond 1000, there is gradual reduction in both the magnitude and frequency of these fluctuations, demonstrating the increasing consistency in the agent's control actions. This continued refinement of the policy yields a more stable closed-loop behaviour.

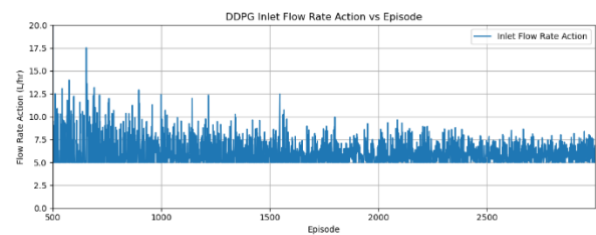


Figure 12: DDPG-trained controller inlet flow-rate control actions across training episodes

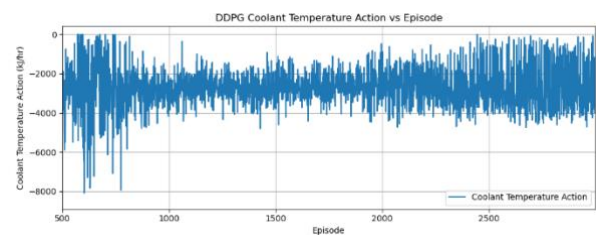


Figure 13: DDPG-trained controller coolant temperature control actions across training episodes

Figures 12 and 13 represent the control action profiles of the two manipulated variables for this CSTR system: inlet flow-rate F and cooling duty $\dot{Q}$. The trends shown in the figures are consistent with the state trajectories and reward profiles in that during the early training episodes there is high variability in the control actions, which is attributed to the DDPG agent's exploration of the system. However, as the number of episodes increases the control actions tend toward a narrower operating region and remains within this band. This reflects the flexibility required to handle the system nonlinearity, as ultimately the state variables do converge to their set-points.

In summary, the results demonstrate that the DDPG agent was able to learn a stable and effective control policy for the nonlinear CSTR system. In the early learning phases, all four reactor states were characterised by large fluctuations and poor tracking performance. After approximately 500 episodes, there was a clear transition in the learning process with all four reactor states converging to within a narrow band of their set-points, thus indicating that the agent had learned reliable closed-loop behaviour. Beyond this point, the stability of the closed-loop continued to increase, due to the continuous refinement of the control policy.

5. Conclusions

This research has highlighted the potential of reinforcement learning (RL) as a viable framework for process control. In the first case study, the A2C-trained controller demonstrated that RL is capable of learning the closed-loop behaviour required for flow control, by successfully tracking the 0.6 and 0.8 L/min set-points. Due to the skewed penalty weightings in the reward function, the controller was unable to converge for the lower set-points, however with refinement of these weightings, it is plausible that the controller could generalise set-point tracking across the full range of tested flow-rates.

In the second case study, the DDPG controller was shown to learn a stable control policy for a nonlinear CSTR exhibiting van de Vusse kinetics. The learning process showed a clear refinement of the control policy from the early episodes, where the trajectories were erratic due to random exploration, to later episodes in which the agent was able to successfully track the state to within a narrow band of the set-point. This progression confirms that the agent ultimately achieved an effective policy for stable closed-loop control within the nonlinear and multivariable reactor system.

Overall, the two case studies explored in this research show that RL can reproduce PID-level closed-loop behaviour for basic flow control applications, while also providing a viable control solution for nonlinear

and multivariable dynamics that challenge PID control systems. These findings demonstrate the broader potential of RL as an emerging solution for advanced process control solutions.

6. Outlook

Further development of both case studies is required to enhance their robustness and progress toward practical implementation. The first case study would largely benefit from optimisation of the penalty weightings, to identify the combinations that best satisfy all components of the reward function and promote more robust control policies. In addition to this, diversifying the training environment of the A2C-based controller to a dynamic transfer function, built from a much wider set of historical process data, would enable a more representative state-action mapping of the system. Therefore, allowing the agent to generalise tracking across a wider range of set-points. This would be highly beneficial for simple processes, like the flow application considered in the first case study, as the A2C-trained controller would not require re-tuning when moving from low, medium and high operating points, as PID controllers would require.

For the second case study, the future work should address the limitations associated with a purely model-based training approach. The current implementation relies entirely on the idealised van de Vusse reactor model, which fails to capture the unmodelled dynamics and disturbances present in real systems. Therefore, training the agent using disturbance terms in the ODE, emulating actuator noise within the state variables or altering the environment from the system of ODEs to an offline pre-collected dataset with broader range of operating conditions can accurately reflect the real system dynamics.

Ultimately, in order to achieve practical deployment of these offline-trained RL controllers, online fine-tuning would still be required to expose the learned policy to the real process dynamics. This would allow the controller to adapt to unmodelled disturbances, plant nonlinearities and process drift, which cannot be fully captured in datasets or simulations. Therefore, online fine-tuning is a necessary step in progressing toward the practical deployment of RL-trained controllers.

Looking ahead, reinforcement learning has the potential to offer several benefits for industrial process control. By enabling RL controllers that adapt autonomously to shifts in operating condition, it has the ability to reduce the supervisory workload on process operators. Furthermore, RL has the ability to incorporate constraint handling directly into the learning framework, offering a significant advantage

over PID control. Explicit constraint handling can prevent common plant issues such as actuator saturation, which is frequently caused by integral windup in PID loops. In this regard, RL offers a potentially more robust framework for future industrial control applications.

Acknowledgements

We would like to sincerely thank Dr Mohamed Maher for his invaluable support, guidance and constructive feedback throughout our research project.

References

- [1] King, M. *Process control: A practical approach*. 2nd edn. Wiley 2016,
- [2] Aström, K. J., & Hägglund, T. (1995). *PID controllers: Theory, design, and tuning* (2nd ed.). Instrument Society of America.
- [3] Seborg, D. E., Edgar, T. F., Mellichamp, D. A., & Doyle, F. J. III. (2016). *Process dynamics and control* (4th ed.). John Wiley & Sons.
- [4] Yahagi, S.; Kajiwar, I. Noniterative Data-Driven Gain-Scheduled Controller Design Based on Fictitious Reference Signal. *IEEE Access* 2023,
- [5] Rastegarpour, S., Feyzmahdavian, H. R., & Isaksson, A. J. (2025). *Adaptive ensemble reinforcement learning for industrial process control*. *Journal of Process Control*, 156, 103575.
- [6] Devarakonda, V.S.; Sun, W.; Tang, X.; Tian, Y. Recent Advances in Reinforcement Learning for Chemical Process Control. *Processes* 2025, 13, 1791
- [7] Luz, E. M. L., & Caarls, W. (n.d.). *Comparison of reinforcement learning techniques for controlling a CSTR process*. Department of Electrical Engineering,
- [8] Bujgoi, G. & Sendrescu, D., 2025. *Tuning of PID Controllers Using Reinforcement Learning for Nonlinear System Control*. *Processes*, 13(3), p.735.
- [9] Bloor, M., Ahmed, A., Kotecha, N., Mercangöz, M., Tsay, C. and del Rio-Chanona, E.A., 2025. Control-informed reinforcement learning for chemical processes. *Industrial & Engineering Chemistry Research*, 64(11), pp.4966–4978
- [10] Wang, H., Kontoravdi, C., & Del Rio Chanona, A. (2025). *Offline reinforcement learning for bioprocess optimization with historical data*. Sargent Centre for Process Systems Engineering, Imperial College London.
- [11] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
- [12] Rajpoot, V. (2020). *Control CSTR using reinforcement learning* (Version X.X) [Source code]. GitHub.
<https://github.com/vikas-rajpoot/Contorl-CSTR-using-Reinforcement-learning>
- [13] Green, D. W., & Southard, M. Z. (Eds.). (2019). *Perry's Chemical Engineers' Handbook* (9th ed., Chapter 10.1: Transport and Storage of Fluids). McGraw-Hill Education.
- [14] Wikipedia contributors. (2025, November 16). *Diaphragm pump*. In *Wikipedia, The Free Encyclopedia*. Retrieved December 4, 2025, from https://en.wikipedia.org/wiki/Diaphragm_pump

Extracting Chitin from Mushrooms with Ionic Liquids

Helen Lee

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

This paper assesses the use of mushrooms as a vegan alternative to the traditional chitin source, crustaceans. First, the components of raw mushroom biomass were characterized using stepwise conventional extraction steps as detailed in literature. This led to underestimation of protein and carbohydrates. Conventional acid hydrolysis and alkaline extraction were used as a baseline to assess effectiveness, with alkaline extraction achieving results consistent with mushroom chitin in literature. The use of an ionic liquid, which are salts that are liquid at room temperature, for chitin extraction was then investigated. A design of experiment (DoE) was created to investigate the effect of temperature, solids loading, and time for the use of acidic ionic liquid DMBA · H₂SO₄ to extract chitin. The data and trend obtained thorough the DoE indicate promising yield and qualitative purity and the extraction process could be further optimized using the parameters considered in this study.

Keywords: Ionic liquids, fungal chitin, chitin extraction, lion’s mane, reishi, mushroom

1 Introduction

Chitin is the second most common polysaccharide in nature, found in the exoskeletons of many crustaceans and insects, and the cell walls of mushrooms. It has a wide variety of uses, from food thickening, to fertilizer, to wound dressings. Historically, the main source of chitin has been crustacean shells. Due to sustainability and cost concerns, plant-based alternatives to crustaceans have been sought, including mushrooms, which have emerged as a vegan alternative source of chitin. Advantages of mushrooms include the possibility to cultivate year-round and rapid growth [30].

There is increasing interest in mushroom biomass as a source of chitin, and several approaches have been taken in literature to obtain that chitin. However, the research area lacks a distinct standard method of isolation, and reported purity and yields show great variability. Commonly employed conventional methods, such as acid hydrolysis, often result in chitin of low purity that contains protein, carbohydrates, and other contaminants.

Ionic liquids (ILs) provide a promising alternative to conventional extraction methods. Often, they are used as solvents to extract biopolymers from biomass sources due to their ability to disrupt hydrogen bonding networks. Ionic liquids have been used to extract chitin, cellulose, and hemicellulose from squid pens and crustaceans. However, there is little research surrounding the use of ionic liquids for chitin extraction from fungal biomass and the effects of parameters on the chitin obtained.

This study aims to (i) characterize the components of raw mushroom biomass, (ii) assess the effectiveness of conventional methods of chitin extraction and (iii) assess

the performance of an ionic liquid for the same purpose.

2 Background

Mushrooms have relatively complex structure. Dried mushrooms are expected to contain:

1. Carbohydrates
2. Chitin
3. Proteins
4. Lipids
5. Minerals

Total carbohydrate content of mushrooms can be expected to be 50-75% of dry weight for most mushroom species, with glucans making up the bulk, notably β -glucans [28]. These form a complex with the chitin in the mushroom, with the overall chitin content expected to be about 10-25% by weight of dried mushroom [29]. Following, protein generally comprises 20-35% by dry weight [22]. Lipid content is generally low in mushrooms, and can be expected to fall somewhere in the 6-15% percent range, with ash (mineral) content being 3-11% percent [6].

There are many methods of characterizing the components of mushroom biomass. Firstly, solvent extraction is employed for lipid extraction. Commonly, ethanol is used as a solvent [11]. Methanol and hexane are also used, but due to increased safety concerns from these solvents, and a worse selection range, ethanol dominates as the solvent of choice [34], [25].

Next, deproteinization is required to assess the protein content of the biomass. The most common methods of deproteinization are acid or alkali hydrolysis. Acid hydrolysis is commonly completed with HCl [27], [12]. Alkali hydrolysis using NaOH is also found throughout literature [16], [12], [15]. These methods are often harsh and denature some amount of the protein, so for the purpose of characterizing the amount of protein, the protein content is often also estimated. Most commonly, the Kjeldahl method is used to relate the total nitrogen content of a material to its protein content using a relevant factor [1]. For mushrooms, this factor is taken as 4.38 over the standard value of 6.45 due to the higher non-protein nitrogen contribution from chitin as compared to other biomasses [10],[23], [20].

Next, the carbohydrate content of the mushroom biomass is extracted. Hot water or dilute acid extraction is often used for the simpler sugars that are soluble [18]. Alkali extraction is required for glucans and strongly bound sugars [32]. As the mushrooms are expected to contain larger amounts of strongly bound β -glucans, the preferred choice is alkali extraction.

Finally, the mineral content of the sample can be determined through ashing. The remaining residue from the final alkali extraction is theoretically chitin.

The purity of the final chitin samples obtained through these methods are often limited without more processing, due to their tendency to create chitin-glucan matrices as opposed to pure chitin. The β -glucans and other polysaccharides contained in the fungal cell wall are tightly linked to the chitin through hydrogen bonding and covalent linkage, which make the mushrooms rigid [18]. As a result, treatment with NaOH often leaves behind alkali-insoluble glucans and protein.

Therefore, use of an ionic liquid for the fractionization of mushroom biomass is investigated as a method of procuring the chitin. The use of an ionic liquid has the great benefit of the potential to preserve the structure of the fungal chitin and have better selection than that of the conventional method. Some studies have reported using [BMIM][HSO₄] or [C₂mim][OAc], but the purity and yield of the reported resulting chitin vary greatly. Some reported values vary from 9–30% yield and 40–70% purity, showing the great variability of the process conditions and chosen IL [26]. Ionic liquids remain a promising pathway for isolating chitin from mushrooms, but optimization of parameters is critical to performance.

3 Methods

3.1 Mushroom Preparation

100 grams of lion’s mane (*hericium erinaceus*) and reishi (*ganoderma lucidum*) mushrooms were obtained from Merit Mushrooms as dried samples. The mushrooms were processed and then ground in a coffee grinder until a powder was achieved, then sieved and separated into particle

sizes. Due to the fibrous nature of reishi mushrooms, the samples were not broken down enough to be sieved, so experiments were performed on one larger particle size.

3.2 Stepwise Compositional Analysis

Following mushroom preparation, the following methods were used to extract various mushroom components.

Soxhlet Extraction Soxhlet extraction is a technique that uses a solvent to extract lipids. Two samples each of reishi, <180 micrometer particle size (small) lion’s mane, and 180-850 micrometer particle size (medium) lion’s mane were soxhlet extracted overnight with absolute ethanol to extract lipids. The lipids were filtered from the extract.

Mild Alkaline Extraction Mild alkaline extraction uses an alkaloid at less intense conditions to extract protein. The defatted biomass was then stirred in 0.1 M NaOH solution with solids loading of 5% for 3 rounds of 30 minutes at 30 C. The solid protein was then precipitated from the resulting liquid drop-wise using 3 M HCl solution, bringing the solutions to pH of 3.5.

Strong Alkaline Extraction Strong alkaline extraction uses an alkaloid at more intense conditions to extract carbohydrates. The deproteinized biomass was freeze-dried and subjected to strong alkaline extraction using 2 M NaOH solution, at 5% solid loading for 2 hrs at 100 C. The carbohydrates were precipitated from the resulting solution using water.

Acid Hydrolysis Acid hydrolysis uses a strong acid, intending to fully hydrolyze any remaining protein and carbohydrates. Finally, the dried biomass underwent acid hydrolysis using 2 M H₂SO₄ solution for 3 hrs at 100 C and continuously stirred. The remaining solids were dried and weighed as the resulting stepwise chitin fraction.

Ash A sample of each raw mushroom was heated in the muffle oven to determine the component of ash.

A flowchart showing the stepwise compositional analysis procedure is shown in Figure 1. Due to the low amounts of ash in biomass, the ash content of the final chitin residue is considered negligible following literature.

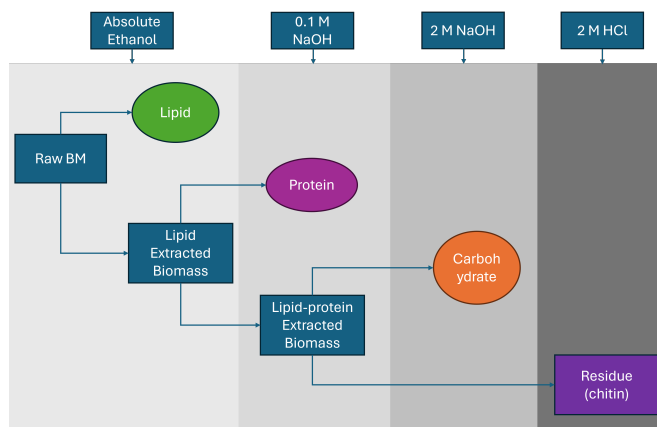


Figure 1: Stepwise Process for Compositional Analysis

3.3 Conventional Method

3.3.1 Alkaline Extraction and Acid Hydrolysis to Obtain Chitin

In this method, defatted biomass underwent alkaline extraction using 2M NaOH solution at 1% solid loading for 2 hours. NaOH is the most commonly used alkaloid in literature due to its success in separating protein, and its promotion of a level of deacetylation that helps to make chitin more accessible to be removed. A middle time of 2 hours was chosen to balance between removal and degradation of the biomass. The used 5% loading during the stepwise compositional analysis formed a thick paste, so the solids loading was lowered to ensure even penetration of liquid. Next, the biomass was contacted with HCl solution overnight at the same solids loading (1%). This was intended to remove the remaining compounds that become acid soluble after the initial treatment.

3.3.2 Direct Acid Hydrolysis to Obtain Chitin

In this method, the soxhlet extracted biomass was directly subjected to overnight acid hydrolysis using H_2SO_4 at 1% solids loading. This is a commonly used technique observed in literature, as it has the ability to penetrate the fungal cell wall and further extract the chitin.

Yield was calculated according to Equation 2 in the supplementary material.

3.4 Factorial Design Optimization

Lion’s mane mushrooms were chosen to be the subject of this design of experiment due to the ease of processing and smaller particle sizes achievable over reishi. Although reishi has a comparatively higher chitin content as reported in literature, its tough structure may make it more difficult to extract. An initial ionic liquid screening was completed with an acidic and alkaline ionic liquid, and the observed yields were lower (better isolation) with the acidic liquid. Therefore, the acidic ionic liquid, $\text{DMBA} \cdot \text{H}_2\text{SO}_4$, was chosen to investigate the effect of solids loading, temperature, and extraction time on the yield of resulting chitin samples. A estimation of chitin purity calculated using the nitrogen content of each sample was also included as a qualitative response variable.

A design of experiment (DoE) was constructed using a full factorial design with 3 center points. The DoE was implemented to simplify the full 27-dataset to an 11-dataset. A summary table of high, low, and center variables is included in Table 1 below, and the full DoE summary table can be found in Table 4.

For each sample numbered 1-11, raw mushroom biomass was contacted with ionic liquid, stirred at 300 rpm using a stir bar and plate, and subjected to the conditions as outlined in the DoE. Following the extraction, each sample was vacuum filtered and washed with water until the filtered liquid appeared clear.

Table 1: Experimental Factor Levels

Factor	High	Center	Low
Temperature ($^{\circ}\text{C}$)	120	100	80
Time (hr)	5	3	1
Solids Loading (%)	5	3	1

3.5 Characterization Methods

Fourier Transform Infrared Spectroscopy (FT-IR) was completed on each of the intermediates of the process using a Cary 630 spectrometer in the analytical lab to obtain a snapshot of the particle interactions in the material. Thermogravimetric analysis (TGA) was used to assess both the raw material and resulting chitin to determine the onset temperature of degradation and percent remaining at 600C, also in the analytical lab. In it, 3mg of each sample was heated to 105 C and left for 10 minutes. The samples were then heated to 600 C at a rate of 20 C/min, with nitrogen used as a purge gas. External carbon-hydrogen-nitrogen (CHN) analysis was completed through Medac Ltd. and was used to assess raw material, resulting chitin, carbohydrates, and protein to assess the purity of each sample. High performance liquid chromatography (HPLC) was completed in the biomass lab in a Shimadzu LC2050 on each ionic liquid sample after extraction. Chromatograms were compared against a set of calibration standards. An example of HPLC data can be found in Table 9 in the supplementary material.

4 Results

4.1 Raw Mushrooms

TGA, FT-IR, and CHN of raw mushroom were completed to assess the composition of the raw material. The TGA plot in Figure 2 reveals that reishi mushroom is a more stable material than lion’s mane. This is attributed to the higher onset temperature for degradation of material. The onset temperature is defined as the temperature at which significant mass loss starts to occur [4]. For reishi, this onset is at 288 C compared to 240 C for lion’s mane, which is consistent with reishi’s more fibrous and denser texture.

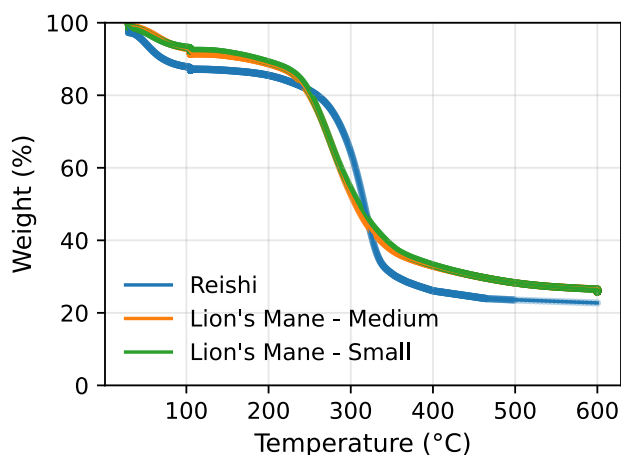


Figure 2: TGA of Raw Mushroom

In FT-IR, a band at 3260 cm^{-1} indicates O-H stretching, which is clearly visible in the FT-IR spectra of the raw mushrooms (Figure 3) [21]. The protein in the mushroom can be observed through the characteristic amide I and II bands at 1630 and 1510 cm^{-1} [5]. Carbohydrates such as β -glucans show strong C-O-C vibrations around 1000 cm^{-1} [33]. Intensity of the glucan-associated bands was stronger in the lion's mane.

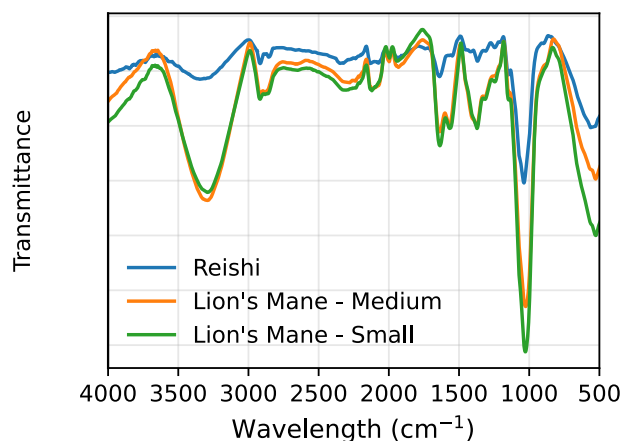


Figure 3: FT-IR of Raw Mushroom

Protein content was estimated from elemental CHN analysis (Table 2) using a nitrogen conversion factor of 4.38 as described in the background. Lion's mane showed higher nitrogen content, and therefore protein content, than reishi mushrooms.

Table 2: CHN of Raw Mushroom

Mushroom	C (%)	H (%)	N (%)	Protein (%)
Reishi	45.09	5.91	1.50	6.52
Lion's Mane (M)	37.53	6.25	2.22	9.67
Lion's Mane (S)	38.28	6.18	2.58	11.23

4.2 Stepwise Characterization of Raw Mushroom Biomass

Analysis of the mass fractions of each step show that lion's mane mushrooms contain $12.5 \pm 2.5\%$ lipid, $32.5 \pm 6.5\%$ carbohydrate, $3.0 \pm 2.7\%$ protein, $27 \pm 7.0\%$ chitin, and $7.5 \pm 2.5\%$ ash. This shows a mass closure of $87.5 \pm 3.0\%$.

Carbohydrate and protein values were underestimated relative to literature and previously estimated values, while the chitin was overestimated. The lipid and ash content aligned with expected values. A full summary table, including values for reishi mushrooms, is included in the supplementary material Table 7.

FT-IR of intermediate solids and products from each step can be viewed in the supplementary material (section 7.3.1). These graphs show confirmation that the expected extractions were achieved at each step. These show the softening of amide bands following protein extraction, and C-O-C bands following carbohydrate extraction as expected.

4.3 Chitin Extracted Using Conventional Extraction Techniques

4.3.1 Yield and Composition

The yield of the chitin obtained from the conventional method of chitin isolation was assessed to determine the effectiveness of both methods, as listed in Table 3.

Table 3: Conventional Method Results

Method	Mushroom	C (%)	H (%)	N (%)	Yield
Acidic	Lion's Mane	45.67	7.06	1.51	0.34
Alkaline	Lion's Mane	46.71	7.20	3.82	0.10
Acidic	Reishi	45.29	6.02	0.79	0.31
Alkaline	Reishi	42.38	6.34	1.01	0.18

As expected following analysis of the raw material, lion's mane mushroom achieves lower yield and therefore better isolation than reishi. The CHN values for reishi (elevated H and lower N) as compared to lion's mane imply protein contamination in the samples. This also manifests in the samples from the acidic method for lion's mane. The alkaline method of extraction appears to be more effective at isolation of chitin due to the lower yield. The CHN analysis of the alkaline method as applied to lion's mane mushroom shows values ($\sim 46\%$ C, 7% H, 3.8% N) that are in line with expected values from similar mushroom chitin samples in literature and is taken as a baseline for the ionic liquid DoE [31].

4.3.2 TGA and FT-IR Analysis

The TGA can be seen in Figure 4, and both curves diverge from expectations. The curve of the alkaline sample is atypically shallow, which suggests unsuccessful TGA. The onset temperature of the acidic conventional method is in line with expected values for glucan-contaminated

mushroom chitin at 260 C. This value, however, is on the lower side of the expected range (240-290 C), which would lead to the conclusion that some degree of deacetylation has occurred [2]. Considering the harsh conditions, it would follow that the resulting chitin at least partially deacetylyzed.

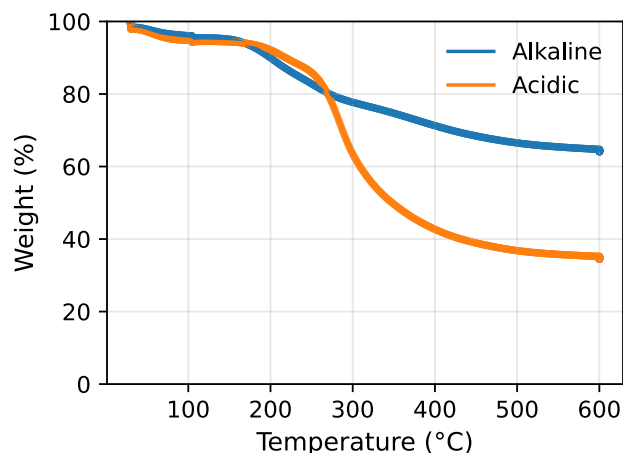


Figure 4: TGA of Conventional Method Chitin from Lion's Mane

The FT-IR in Figure 5 still shows bands that indicate the presence of protein and carbohydrates in the chitin obtained through acid hydrolysis only. The bands are lesser in the chitin obtained through alkaline methods, which follows the CHN analysis and conclusion that the alkaline conditions isolated purer chitin than the acidic method, which observed more significant protein and carbohydrate contamination. However, the spectra for the acidic sample displays distortion and may represent issues with the scan or baseline.

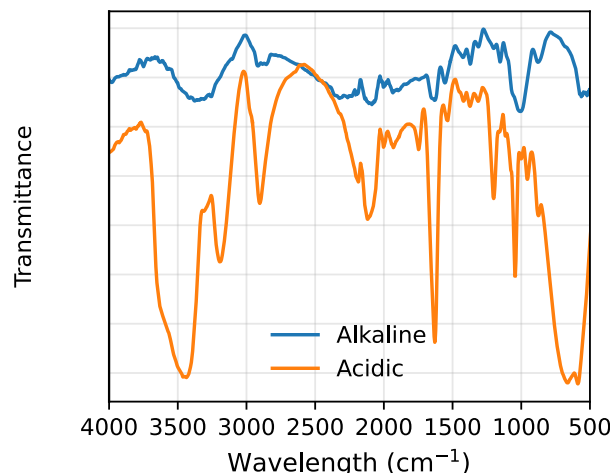


Figure 5: FT-IR of Conventional Method Chitin

4.4 Chitin Extracted Using Ionic Liquid

The resulting chitin obtained from the ionic liquid DoE was similarly assessed for chitin yield and observed for qualitative purity.

4.4.1 Yield and Composition

The purity of each sample was estimated based on the nitrogen content as compared to the ideal conventional sample as outlined in Equation 3 in the supplementary material. Any nitrogen percentages that are above the ideal value of 3.82% are considered to have residual IL and/or protein and carbohydrate in the sample.

As can be seen in Table 4, the sample with the lowest yield also had the highest estimated purity. The extreme high conditions are highlighted in green (Sample 10), center in blue (Samples 5-7), and low in orange (Sample 3).

4.4.2 FT-IR and TGA Analysis

The FT-IR of the center and low samples as seen in Figure 6 appear similar to the ideal chitin sample obtained through the alkaline method, with bands that indicate protein and carbohydrates but remain small. The high sample appears similar to the chitin obtained through the acidic conventional method, which is also impure. The observed sharp peaks in the high sample and the alkaline conventional method (Figure 5) are likely caused by increased crystallization or a more homogeneous material. Although there wasn't a significant visual difference between the samples in the conventional method or the ionic liquid DoE (Figure 16), the FT-IR could imply a level of deacetylation or higher crystallinity. The removal of acetyl groups would allow the chitin chains to pack closer, increasing homogeneity of the material and leading to sharper peaks.

The onset temperature as viewed in the TGA in Figure 7 is consistent between the low and center samples, with a value of approximately 245 C. This value higher for the high sample, at approximately 290 C, Both values fall within the expected range.

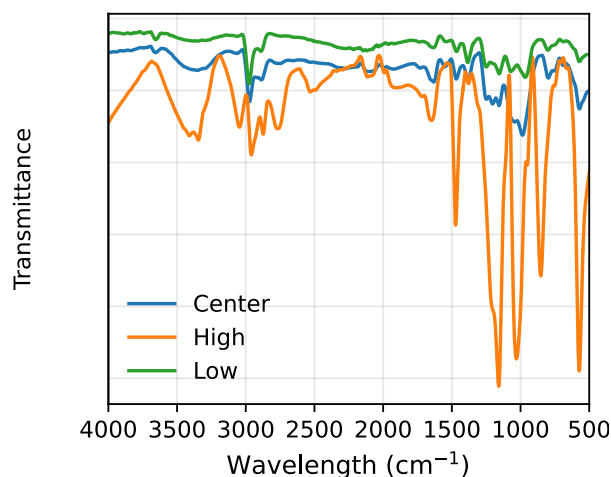


Table 4: Full results of IL DoE

Sample	Temperature	Time	Loading	C	H	N	Yield	Purity Est.
1	120C	1hr	1%	35.61	8.69	6.275	0.69	0.36
2	120C	1hr	5%	38.635	7.705	5.025	0.89	0.68
3	80C	1hr	5%	43.455	7.03	3.865	0.34	0.99
4	80C	1hr	1%	36.905	7.96	5.45	0.88	0.573
5	100C	3hr	3%	36.485	8.42	5.905	0.71	0.45
6	100C	3hr	3%	34.98	8.725	6.22	0.85	0.37
7	100C	3hr	3%	35.835	8.44	5.815	0.59	0.48
8	80C	5hr	5%	38.935	7.665	4.355	0.92	0.86
9	120C	5hr	5%	34.965	8.11	5.42	0.61	0.58
10	120C	5hr	1%	36.145	8.73	6.47	0.59	0.31
11	80C	5hr	1%	33.135	8.685	5.635	0.57	0.52

Figure 6: FT-IR of IL Extreme Points

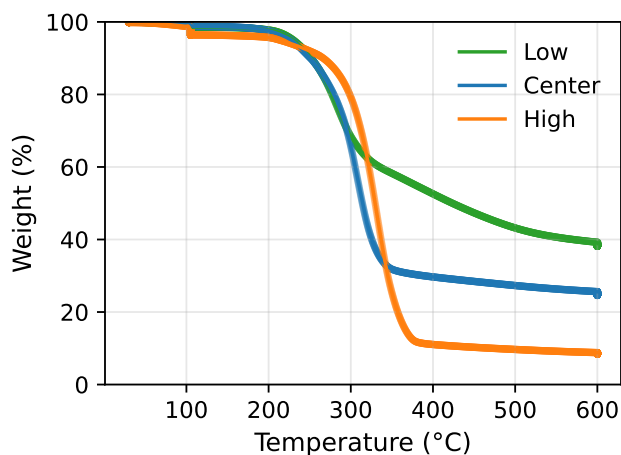


Figure 7: TGA of IL Extreme Points

Some unique issues with processing were observed. The center samples had issues with gelation not observed in the other samples.

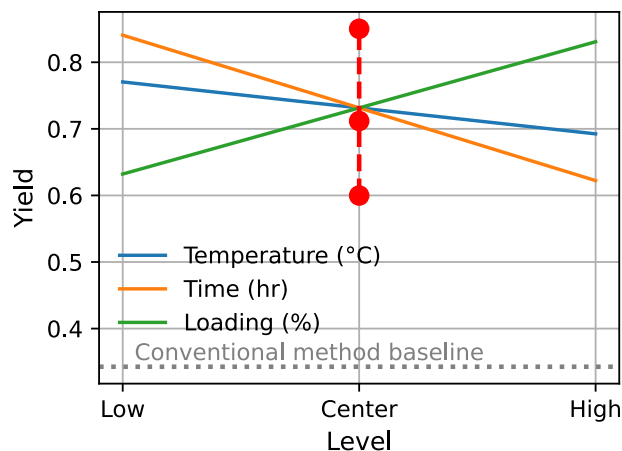


Figure 8: Yield vs. Factor Level

4.4.3 Factor Effects

The effect of the factor levels on the yield are represented in Figures 9 and 8. These curves were obtained by fitting a 4D polynomial surface to the experimental data in a normalized space. The trend was then obtained by taking a slice of the surface. The red points indicate the experimentally determined center points. The surface plots can be found in the supplementary material Figures 17-19. General trends are observed: better isolation (lower yield) occurs with lower loading, higher timer, and temperature. These conclusions follow expectations, as harsher conditions should lead to better extraction. None of the factors at the defined levels approach the baseline for the conventional method, though. It follows that the ionic liquid did not separate chitin as effectively as desired. As can be observed from the slope of the lines in 8 and the bars in 9, time and loading have the greatest effect on yield, and temperature does not effect the yield as greatly, although none prove statistically significant. This implies process variability.

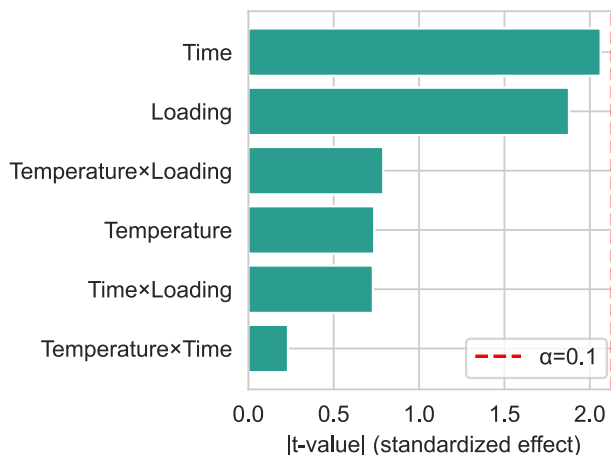


Figure 9: Pareto Chart for Yield

4.4.4 Post Extraction Ionic Liquid Analysis

Finally, HPLC was performed on the resulting ionic liquid used for extraction. This detected no glucose remaining in any IL samples post extraction. An example of HPLC data obtained through the samples can be found in supplementary material Table 9.

5 Discussion and Future Work

This study compared conventional chemical methods for chitin isolation from reishi and lion's mane mushrooms and ionic liquid processing for lion's mane mushrooms. The results highlight structural differences between the mushroom types and the advantages and limitations of extraction using acidic ionic liquid DMBA · H₂SO₄.

5.1 Compositional Analysis of Raw Material

The stepwise compositional analysis achieved a mass closure of $87.5 \pm 3.0\%$ and underestimated the contributions of protein and carbohydrates. Several factors contribute.

The carbohydrate composition of mushroom biomass differs from lignocellulosic biomass. Other plant materials contain readily hydrolyzable sugars such as cellulose, whereas mushrooms contains very little hydrolyzable sugar. They instead contain β -glucans, which do not hydrolyze effectively under the alkaline conditions used during the stepwise compositional analysis. This leads to an underestimation of polysaccharide content of the biomass.

A significant portion of the protein content of the mushroom is also lost in the stepwise analysis. Although isolated fungal chitin often contains residual protein, the bulk of the lost protein is lost in the alkaline deproteinization step. The treatment likely caused the proteins to denature into smaller peptides that are lost with the supernatant. Consequently, the protein content as determined from this step is lower than both expected values from literature and the estimated protein content of the raw material. An additional salting out step could be investigated to recover remaining protein fragments, or a different pH shift such as 4-5 could be investigated to recover more of the protein [8].

Additionally, other methods could provide better quantification of the chitin content of the raw material. Bypassing hydrolysis, solid-state nuclear magnetic resonance (NMR) is an alternative spectroscopy measurement that could characterize the material that could quantify chitin in the fungal cell wall without disrupting the native cell wall.

5.2 Extraction of chitin using the conventional method

While the yield of the conventional method was inline with reported chitin content of raw mushrooms, some

values from the CHN analysis and TGA pointed to inconsistencies in the results. Mushroom cell walls contain a chitin-glucan complex that is very tight and difficult to penetrate. The conventional method is not effective in removing β -glucans and proteins, which remain residual. All factors suggest the existence of a chitin-glucan complex as opposed to pure chitin for the samples obtained.

Variability in particle size, specifically for reishi, likely contributed to inconsistent extraction. Improved preparation to process the reishi into smaller, more consistent particles would be important to fully homogenize the biomass and enhance reproducibility.

Complete glucan-chitin disruption would also be desired to remove pure chitin. In literature, more advanced techniques such as enzymatic digestion or TEMPO-mediated glucan oxidation have been employed to fully disrupt the hydrogen and covalent bonding that exists in the glucan-chitin network [3], [19]. Use of these techniques could reduce contamination and improve purity. These methods could provide a better baseline.

5.3 Extraction of chitin using IL

Overall, the use of ionic liquid DMBA · H₂SO₄ for extracting chitin from mushroom biomass resulted in comparable purity and yield of previously reported samples, emerging as a potentially promising mechanism for chitin extraction. However, some suggestions could be made to improve the effectiveness of this the ionic liquid mechanism.

Firstly, the washing and filtering of the final chitin solids could be improved. The resulting chitin samples visually appeared slightly shiny and were flexible. This, along with the slightly elevated nitrogen levels observed from the CHN analysis, point to residual ionic liquid in the chitin. As the samples were filtered with 60-100 mL of water, this leads to the conclusion that the solids require greater volumes of water to more effectively wash the ionic liquid out of the solids. Successful approaches using water as a washing agent in literature see up to 14 cycles of washing [14]. Other approaches include using acetone or other solvents to complete subsequent washes [7].

This presents potential challenges for scale-up of the process. The recovery of expensive ionic liquid would require efficient washing and recycling steps. Although ionic liquid processing is potentially a greener process than traditional acid and base extraction, the processing required to recycle the IL could introduce environmental and economic constraints.

The ionic liquid and parameter choices are also critical to the results. From these results, DMBA · H₂SO₄ doesn't seem to be aggressive enough to fully disrupt the chitin-glucan matrix. Other ionic liquids have shown potential efficacy, including alkaline ionic liquid [C₂mim][OAc]. Optimization of the chosen ionic liquid and parameters for extraction could improve the purity of the obtained

chitin.

Following this, chitin obtained through IL extraction has been shown to often be protein-contaminated. A deproteinization step following the ionic liquid step may need to be employed to fully remove the contaminating protein from the chitin and to fully purify [14].

5.4 Recovering proteins and sugars from the IL

Recovering the proteins and sugars from the ionic liquid are desirable as a by-product for the process.

The use of water as an anti-solvent was assessed, however was unsuccessful in precipitating the protein from the post-extraction ionic liquid. Due to protein's sensitivity to changes in pH, a more precise pH shift could be investigated for the recovery of protein from the ionic liquid. This approach was used successfully in the initial characterization and could be employed with the acidic ionic liquid. The major disadvantage of this approach is that the strong bases that would be required for the pH shift are harmful to the environment, and would be required in large amounts at scale. These bases also have the potential to damage the ILs and reduce IL recyclability.

Salting out of proteins is also a potential method for protein precipitation. The mechanism reduces protein solubility through the addition of large amounts of salts. This method could be effective at a lab scale, but in scaling up would likely cause problems with the purity of precipitated protein. It would also be costly and an environmental concern to deal with the large amount of salt that would be required for the process. Both of these methods should be assessed for economic feasibility, quality of protein obtained, and efficacy.

In addition to providing a byproduct for the process, the recovery of sugar is critical to the recyclability of the ionic liquid. Sugar buildup in the ionic liquid would degrade the efficiency of the extraction. HPLC revealed that no glucose remained in the post-extraction IL, which would suggest that polysaccharide oligomer chains have not been broken down into monomers. For the recovery of sugars, a number of different techniques have been employed in literature. An aqueous alkaline solution could be used to form a bi-phasic system from which oligomer chains can be separated [24]. These can further be processed into the glucose monomer. Similarly to salting out for proteins, this method would require large amounts of anti-solvent that could affect the IL in terms of recycling. Membrane filtration is also a potential method for trapping the soluble sugars from the IL. This method has successfully filtered oligomer chains from ionic liquids. Despite the high costs that are associated with membrane filtration equipment, this method could prove superior as there isn't a need for anti-solvent.

5.5 Mushroom Waste as a Source

This study shows clear rationale for mushrooms as a chitin biomass source. Further, mushroom waste represents an attractive alternative feedstock to raw mushrooms. This improves the sustainability and circularity of the process. Additionally, mushroom waste is richer in chitin due to larger proportions of stems and tougher mushroom parts, which could improve the available chitin proportion of material. Furthermore, the waste has already been mechanically processed, which could work as pretreatment and allow easier extraction. Future work could evaluate the performance of ionic liquids for this feedstock and could assess the economic feasibility and scalability of the process.

6 Conclusions

This paper presents an approach to isolating fungal chitin through the use of an acidic ionic liquid, $\text{DMBA} \cdot \text{H}_2\text{SO}_4$. Promising results were viewed through the results of the paper, expanding upon values in literature for chitin obtained from ionic liquids. The chitin obtained through the use of ionic liquid in this study shows purity and yield that is inline with literature and provides a good baseline. This shows that mushrooms could certainly be a more sustainable, plant-based source of chitin, but the area requires significant work to make the process feasible economically. Further optimization of the process for ionic liquids could potentially provide a step towards higher quality fungal chitin. Overall, ionic liquids establish a tunable and potentially more effective pathway to extract fungal chitin than the conventional method, but effective disruption of the cell wall glucan-chitin structure remains a primary area for future work.

References

- [1] Kjeldahl method – an overview. ScienceDirect Topics, —.
- [2] D. Araújo, V. D. Alves, A. C. Marques, E. Fortunato, M. A. M. Reis, and F. Freitas. Low temperature dissolution of yeast chitin–glucan complex and characterization of the regenerated polymer. *Bioengineering (MDPI)*, 7:28, 2019.
- [3] Sarah M. Bowman and Stephen J. Free. The fungal cell wall: Structure, biosynthesis, and function. *Microbiology and Molecular Biology Reviews*, 70(3):583–604, 2006.
- [4] Malcolm E. Brown and Patrick K. Gallagher. *Handbook of Thermal Analysis and Calorimetry: Principles and Practice*. Elsevier, Amsterdam, 2003.

- [5] W. Chen and H. Zhao. Ftir analysis of mushroom proteins and their structural features. *Food Chemistry*, 272:451–458, 2019.
- [6] A. Dimopoulou and E. et al. Galani. Nutritional composition and biological properties of mushrooms: A review of the health aspects. *Applied Sciences*, 12(16):8074, 2022.
- [7] V. M. Dinh and S. G. et al. Khokarale. Ionic liquid strategy for chitosan production from chitin and molecular insights. *RSC Sustainability*, 2(4):1154–1164, 2024.
- [8] K. C. Duong-Ly and S. B. Gabelli. Salting out: A simple method for protein precipitation. *Methods in Enzymology*, 541:85–94, 2014.
- [9] D. Elieh-Ali-Komi and Others. Chitin and chitosan — structural, compositional and thermal properties: review and baseline data. *Marine Drugs*, 18:64, 2020.
- [10] FAO/INFOODS. Fao/infoods guidelines for conversion among different food data units, 2012.
- [11] B. A. Hanson and R. L. et al. Lester. The extraction of inositol-containing phospholipids and phosphatidylcholine from *saccharomyces cerevisiae* and *neurospora crassa*. *Journal of Lipid Research*, 21(2):309–315, 1980.
- [12] H. K. et al. Ketemepi. Extraction process and applications of mushroom-derived protein hydrolysates: A comprehensive review. *Journal of Food Bioengineering*, 2024.
- [13] X. Kong and Others. Environmental sustainability and physicochemical property screening of chitin and chitin-glucan from 22 fungal species. *Polymers*, 17:2722, 2024.
- [14] D. R. Lyon and B. R. et al. Smith. Deproteinization of chitin extracted with the help of ionic liquids. *Molecules*, 27(13):3983, 2022.
- [15] J. et al. Nitschke. A new method for the quantification of chitin and chitosan. *Carbohydrate Polymers*, 83(4):1573–1578, 2011.
- [16] S. P. Ospina Álvarez and D. A. et al. Ramírez Cadavid. Comparison of extraction methods of chitin from *ganoderma lucidum* mushroom obtained in submerged culture. *BioMed Research International*, 2014:1–6, 2014.
- [17] Kurt S. Pregitzer and Andrew J. Burton. Chn analysis: A reliable method for the determination of carbon, nitrogen, and hydrogen in organic materials. *Analytical Chemistry*, 80(13):4605–4611, 2008.
- [18] F. A. Ruthes and A. et al. Smiderle. Mushroom heteropolysaccharides: A review on extraction, isolation, and structural characterization. *Carbohydrate Polymers*, 136:358–375, 2016.
- [19] Tsuguyuki Saito, Yoshiharu Nishiyama, Jean-Luc Putaux, Michel Vignon, and Akira Isogai. Cellulose nanofibers prepared by tempo-mediated oxidation of native cellulose. *Biomacromolecules*, 7(6):1687–1691, 2006.
- [20] L. Schütte et al. The nitrogen content in the fruiting body and mycelium of fungi. *Microorganisms*, 12(2):431, 2024.
- [21] J. Smith and A. Lee. Infrared spectroscopic characterization of polysaccharides. *Carbohydrate Research*, 346:123–130, 2011.
- [22] N. Stilinović and I. et al. Čapo. Chemical composition, nutritional profile and in vivo antioxidant properties of the cultivated mushroom *coprinus comatus*. *Royal Society Open Science*, 7(9), 2020.
- [23] M. Stojek et al. Protein content and amino acid profile of wild mushrooms depend on species and growth site. *Journal of Functional Foods*, 109:105929, 2025.
- [24] N. Sun and H. et al. Liu. Production and extraction of sugars from switchgrass hydrolyzed in ionic liquids. *Biotechnology for Biofuels*, 6(1):39, 2013.
- [25] M. Tangaraju and Z. et al. Ilham. Fueling the future with fungi: Efficient lipid extraction from *ganoderma* mycelium biomass. *Progress in Energy and Environment*, 29:16–25, 2024.
- [26] V. K. Thakur and S. I. et al. Voicu. Recent advances in ionic liquid-assisted extraction of biopolymers. *Materials Today Chemistry*, 12:282–295, 2019.
- [27] J. Vetter et al. Chitin content of cultivated mushrooms *agaricus bisporus*, *pleurotus ostreatus* and *lentinula edodes*. *Food Chemistry*, 104(1):386–390, 2007.
- [28] J. Vetter et al. The mushroom glucans: Molecules of high biological and medicinal importance. *Foods*, 12(5):1009, 2023.
- [29] T. Wu and S. et al. Zivanovic. Chitin and chitosan: Value-added products from mushroom waste. *Journal of Agricultural and Food Chemistry*, 52(26):7905–7910, 2004.
- [30] A. Zannat and Others. Comparative analysis of chitin isolation techniques from mushrooms: toward sustainable production of high-purity biopolymer. *Green Chemistry*, 2025.

- [31] J. Zhang and W. et al. Liu. Extraction and characterization of chitin from mushroom waste. *International Journal of Biological Macromolecules*, 123:1000–1008, 2019.
- [32] M. Zhang and H. C. et al. de Bruyn. Extraction and characterization of polysaccharides from edible mushrooms by different methods. *Food Hydrocolloids*, 25(5):1028–1035, 2011.
- [33] Y. Zhang and X. Wang. Infrared characterization of β -glucans in fungal biomass. *Journal of Agricultural and Food Chemistry*, 65:1203–1210, 2017.
- [34] M. Čertík and P. et al. Andráš. Effect of extraction methods on lipid yield and fatty acid composition of lipid classes containing α -linolenic acid extracted from fungi. *Journal of the American Oil Chemists' Society*, 73(4):357–365, 1996.

Cost Optimisation of Hydrogen Integration in the ASEAN Power Grid Using TZ-APG PyPSA Model

Tak Fai Lam & Justin Too Kah Fung

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Hydrogen is a promising vector for decarbonisation, especially in ASEAN, where power systems are heavily dominated by fossil fuels. This report investigates how hydrogen could be integrated into ASEAN's regional power grid at minimum cost, using Transition Zero's Two-Zone ASEAN Power Grid (TZ-APG) model implemented in PyPSA. The model can co-optimize generation, transmission, and hydrogen conversion across all 10 ASEAN countries, represented by 24 buses, over a full year at 2-hour resolution. Three scenarios were examined: the 2023 baseline, the 2035 no-hydrogen scenario, and the 2035 hydrogen scenarios. Results show that introducing hydrogen in 2035 reshapes the generation mix by generally reducing hydropower while increasing biomass and, to a lesser extent, coal and gas. The total annual system cost rose from €111.75 billion to €145.41 billion, mainly driven by the cost of electrolyser, hydrogen transport infrastructure (hydrogen links), and additional renewable investments. The implied levelized cost of hydrogen in the 2035 scenario is about €4.2/kg (~USD \$5/kg), which is higher than the typical global targets for hydrogen and indicates that hydrogen production remains a relatively costly option under the current assumptions. Thus, ASEAN will need substantial enabling infrastructure and coordination to meet its policy-driven national goals.

Keywords: Green Hydrogen, ASEAN Power Grid, PyPSA, TZ-APG model, Energy System Co-Optimisation

1. Introduction

The Association of Southeast Asian Nations (ASEAN) has become one of the world's fastest-growing regions, driven by rapid economic expansion, population growth, and urbanisation. However, ASEAN still faces a major challenge in meeting its growing energy demand as the region transitions toward a low-carbon economy. Many ASEAN countries still rely heavily on coal, oil, and natural gas, which account for over 75% of the region's primary energy supply. [1] Whilst Organisation for Economic Co-operation and Development (OECD) countries have made significant progress in reducing per capita emissions to 7.9 tonnes in 2021 since mid-2000, [2] developing Asia still faces a challenging task of balancing economic growth, energy affordability, and environmental protection. Thus, to achieve a green and sustainable transition, ASEAN countries must diversify their energy mix by scaling up renewables and integrating clean energy carriers such as hydrogen into their regional system.

Globally, about 120 million tonnes of hydrogen are produced annually, of which approximately 95% of it is derived from fossil fuel without carbon capture ("grey hydrogen"), and only a small fraction of that comes from renewable sources ("green hydrogen"). [3] However, this trend is changing rapidly, as reducing greenhouse gas emissions is a high priority under COP21 and the upcoming COP30, which require leaders to pursue alternative fuel pathways, thereby shifting away from fossil fuels. The ASEAN Low-Carbon Energy Technologies Roadmap (ALERT) has identified hydrogen as one of the five strategic technologies essential to achieving carbon neutrality by 2050.

Currently, widespread hydrogen adoption remains limited, constrained by high production costs, limited

refuelling infrastructure, and the lack of a system-level framework linking electricity generation with demand. Thus, to overcome these challenges, hydrogen must be integrated into energy system planning, enabling coordinated optimisation across generation, transmission, and infrastructure. To analyse these variables, this project builds on Transition Zero's Two-Zone ASEAN Power Grid (TZ-APG) model, an open-source Python-based Power System Analysis (PyPSA) model. The TZ-APG model can simulate electricity generation, storage, and cross-border transmission across all 10 ASEAN countries within a least-cost optimisation framework. [4] This project extends TZ-APG by integrating hydrogen electrolysis, and transport demand, enabling hydrogen to serve as an additional carrier within the system.

Therefore, this project aims to evaluate the least-cost integration of hydrogen within ASEAN's Power Grid using the TZ-APG PyPSA framework. This study will show how integrating hydrogen production, storage, and transport demand will affect the region's energy mix and total system cost.

2. Literature Review

2.1 ASEAN's Energy Landscape

The ASEAN region has long recognised the need for a deeper, more integrated energy system to reduce reliance on fossil fuel and support large-scale renewable energy projects. The ASEAN Interconnection Masterplan Study (AIMS) III is the most comprehensive planning report, detailing 18 priority interconnection projects across the region to achieve a combined capacity target of up to 33GW by 2040. [5] AIMS III has demonstrated that regional interconnection can reduce costs, optimise the utilisation of diverse renewable resources, and mitigate reliance on coal and gas generation. However, despite

nearly three decades of regional planning efforts, the region's grid has largely remained fragmented, with cross-border lines predominantly unidirectional and bilateral, thereby limiting the extent to which renewable energy can be shared across the region. This fragmented energy system places pressure on ASEAN countries to expand variable renewable generations and to develop complementary options such as hydrogen-based systems that is able to provide long-duration storage and support transmission expansion. Yet neither AIMS III nor any existing regional modelling frameworks have integrated hydrogen technologies, leaving an important gap in understanding how hydrogen might play a role in the long-term grid development, even though several countries are already exploring hydrogen as an energy vector in their national plans.

2.2 Hydrogen as a Flexibility Technology

Hydrogen has emerged as a promising and versatile energy carrier with significant potential to decarbonize power systems worldwide. In the power sector, power-to-hydrogen-to-power (P2H2P) provides a long-duration buffer for renewable variability over days, weeks, and seasons. Unlike batteries, which are only cost-effective for short-term hourly fluctuations, hydrogen storage can provide large energy capacity at relatively low cost, especially in underground salt caverns. For instance, the UK Energy Storage (UKEn) has developed hydrogen storage projects in salt caverns, which would lead to a significant lower cost per kg of hydrogen stored due to its lower operating cost. [6] The European energy planning system in 2050 showed that coupling hydrogen storage with grid expansion can reduce electrical storage requirements by 35% and replace 40% of electrical transmission capacity. [7] However, the deployment of hydrogen infrastructure is highly sensitive to techno-economic factors. Three factors in particular dominate the cost-effectiveness of hydrogen production: electrolyser capital cost (CAPEX), conversion efficiency throughout the hydrogen chain, and renewable availability.

The Economic Research Institute of ASEAN and East Asia (ERIA) has shown that hydrogen systems are currently more expensive than other flexibility options. Nonetheless, they become more cost-effective in power systems with high renewable penetration and stringent carbon limits. Furthermore, by 2040, the cost of hydrogen would decrease by 50% if it is adopted across all sectors, with a target price of \$2.00-\$2.50/kg, driven by advances in electrolyser technologies and economies of scale. [8] Although round-trip efficiencies for hydrogen electrolyser systems are typically 30-40%, they can still economically store massive amounts of energy that batteries cannot, thus bridging periods of low renewable output.

2.3 Hydrogen Demand in ASEAN

Currently, ASEAN's hydrogen demand is roughly at 3.7 million tons per year, and it is supplied almost entirely via natural gas. Indonesia and Thailand are major players in

the refining sector, with the capacity to produce up to 1.13 and 1.10 million barrels in Thailand and Indonesia, respectively, accounting for the majority of ASEAN's hydrogen demand. [9] Projections indicate that ASEAN will have growing hydrogen demand, though this may vary by scenario and sector, as shown in *Figure 1*. Transportation and industry will most likely dominate hydrogen demand, while electricity generation will account for only a small share. [10] Furthermore, if hydrogen is blended at 10% with natural gas, ASEAN's hydrogen demand could reach up to 6.6 million tons per year by 2040. This led to a significant reduction in carbon dioxide emissions and an increase in hydrogen's share of electricity generation from 10% to 30%. [9]

Across ASEAN, four countries have published their national hydrogen strategies as of 2022-2024, including Singapore, Malaysia, Indonesia, and Vietnam. All four countries aim to capitalise on the global momentum surrounding hydrogen by emphasising a phased-based approach across different time horizons to achieve decarbonization targets. Countries like Malaysia, Indonesia, and Vietnam, with their abundant natural resources, have also incorporated hydrogen export-import strategies into their roadmaps. [11]

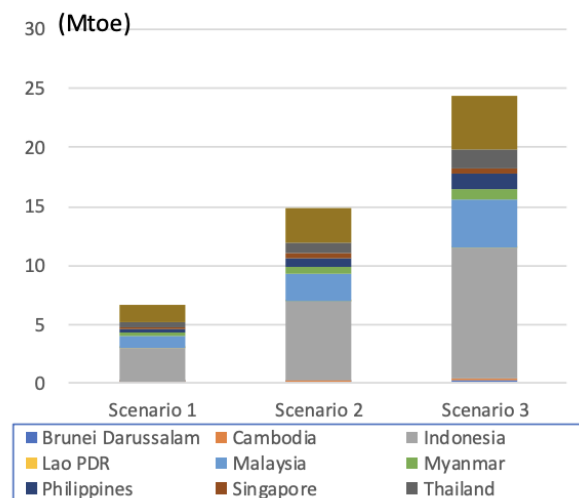


Figure 1: Hydrogen Demand by Country under Various Scenarios by 2040

2.4 Current Developments of Hydrogen Integration in Power Systems

There has been rapid progress in modelling hydrogen integration within power and energy systems, especially in Europe and other OECD regions. A recent work using a PyPSA-based model compared electricity-grid expansion with a dedicated hydrogen pipeline network and found that a continental hydrogen grid can lower total system costs and partially substitute for new transmission capacity by transporting energy as “molecules rather than electrons”. [12] At the same time, a report has shown that many hydrogen studies have omitted various key features such as sector coupling, temporal variability, and endogenous learning in the sector-coupled European

model, thus leading to a much faster and cost-optimal scale-up of electrolyzers and renewables compared to what is envisaged by the EU, reducing hydrogen cost to around €1.26/kg. [13]

Within ASEAN, energy-system modelling has already begun to explore hydrogen as a means to absorb curtailed renewable generation and reduce dependence on fossil fuels. While the region is abundant in solar and wind resources, they contribute only marginally due to weak grid integration. Hence, studies have been conducted to evaluate scenarios in which curtailed wind and solar power would be diverted to electrolyzers, producing hydrogen that can be stored and transported for multiple end uses. [14] Hence, to address these gaps in the literature, this project transfers established European and OECD modelling approaches to the ASEAN context. We extend the TZ-APG PyPSA model by adding electrolyzers, country-level hydrogen demand and technology assumptions. The next section sets out this methodology, along with the scenarios used to assess how hydrogen could influence ASEAN's future grid development.

3. Methodology

3.1 PyPSA Modelling

This model is conducted using PyPSA, an open-source optimisation model designed for large-scale energy systems. PyPSA provides a mathematical representation of multi-carrier networks and supports the co-optimisation of generation, technology conversion, and transmission across high-resolution temporal datasets. PyPSA can simultaneously optimise investment and dispatch decisions, enabling it to identify the least-cost infrastructure to satisfy electricity and hydrogen demand across various scenarios.

The underlying electricity network is derived via TZ-APG's model, which provides a representation of the power systems across all 10 ASEAN countries. TZ-APG maps the region into 24 different buses, each corresponding to a major load centre, renewable resource area, or national balancing zone. The model also incorporates existing cross-border transmission, enabling power exchange within the region. TZ-APG also provides a detailed generator portfolio, renewable availability profiles, and hourly demand data, allowing the model to reflect real-world operation. Furthermore, for this project, the model is run for a full year at a two-hour resolution, providing 4,380 snapshots per scenario. This timestep is sufficient to represent a daily renewable variability without compromising any computational efficiency.

3.2 Business-as-Usual Framework

The Business-as-Usual (BAU) configuration serves as the reference configuration for both the baseline scenario 2023 and future scenarios, as shown in *Figure 2*. By fixing the transmission topology, the BAU structure

provides a stable, policy-consistent baseline for assessing hydrogen integration.

3.3 Scenarios

Three scenarios have been developed for this study:

The 2023 Baseline Scenario represents the current ASEAN power system using data directly from the TZ-APG model. It includes existing generation fleets, present-day transmission links, and renewable availability profiles. This scenario provides a reference point against which future system developments can be compared and validated.

The 2035 No-Hydrogen Scenario represents the future power system under BAU conditions without the introduction of hydrogen production or hydrogen-related technologies. Electricity demand is scaled to 2035 levels, renewable technologies are allowed to expand endogenously wherever cost-effective, while fossil generation capacities remain fixed at 2023 levels.

The 2035 Hydrogen Scenario also builds upon the BAU 2035 system, but it introduces a full power-to-hydrogen subsystem. Technologies such as electrolyzers, fuel cells, and dedicated hydrogen buses for each country would be added, along with an exogenously specified hydrogen demand profile.

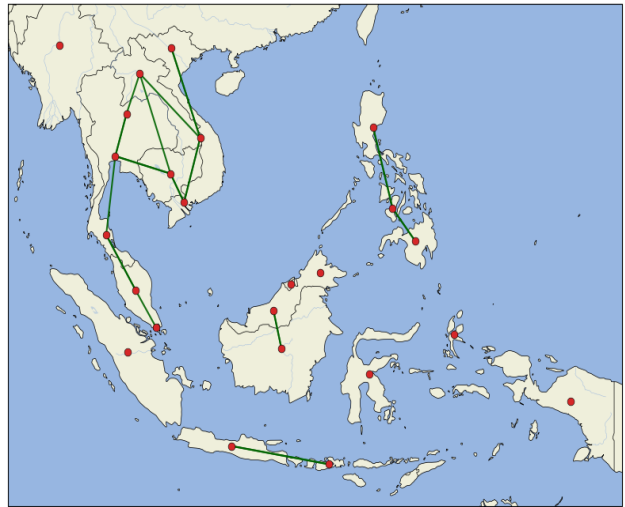


Figure 2: Business-As-Usual existing transmissions

3.4 Basic Infrastructure Design

In each country, a dedicated new hydrogen energy bus is connected to its national electricity buses. Conversion technologies, such as electrolyzers that convert electricity to hydrogen, were modelled as controllable links in PyPSA. A fixed hydrogen demand was imposed to certain countries based on literature, shown in Table 1. [15] [16] [17] [18] [19] Specifically, certain countries would be assigned to a targeted volume of hydrogen that has to be produced within this model to ensure that hydrogen is able to account for certain fractions of generation mix by 2035. Country-level hydrogen demands were

interpolated to estimate the hydrogen demand for 2035. Countries without an official hydrogen strategy were excluded from the analysis, as incorporating unsupported documented estimates would introduce speculative bias. Within PyPSA, this is implemented as a must-serve hydrogen load profile on the hydrogen bus, ensuring that the model produces and consumes the requisite hydrogen for each period.

Hydrogen demand given in million tonnes is converted into energy units using hydrogen's lower heating value (LHV), where 1kg of $H_2 \approx 33.33kWh$.

Table 1: Current and future projections of H_2 demand in ASEAN
MT: Million Tonnes of Hydrogen

Countries	Current H_2 Demand/year in 2023	Future H_2 Demand/year
Indonesia	1.75 MT	4.2 MT by 2060
Malaysia	0.05 MT	2 MT by 2030
Philippines	0.05 MT	0.2 MT by 2040
Singapore	0.48 MT	0.6 MT by 2040
Thailand	0.32 MT	0.4 MT by 2040
Vietnam	0.60 MT	3MT by 2040

3.5 Electrolyser Technology and Assumptions

Electrolysers are modelled as Proton Exchange Membrane (PEM) electrolysers, which is consistent with the technological pathway expected to dominate green hydrogen production in 2030-2035. [20] PEM electrolysers are selected due to their high operational flexibility and strong compatibility with electrical systems dominated by variable renewable energy (VRE). These characteristics make them appropriate for ASEAN's future power grid, where renewable sources such as hydropower exhibit seasonal variability. Alkaline electrolysers are commercially more mature and typically cheaper in terms of CAPEX; however, they offer slower ramp rates and narrower part-load operation, making them less well-suited to a highly variable power supply. [21] Solid oxide electrolysers have higher electrical efficiency at higher temperatures, but they incur extremely high capital costs and are less commercially ready. [20]

Within PyPSA, electrolysers are modelled as energy conversion links and the conversion process is modelled via fixed efficiency, such that the hydrogen output $p_{ely,t}^{H_2}$ at snapshot t is proportional to the electrical input $p_{ely,t}^E$ shown in Equation 3.5.1

$$p_{ely,t}^{H_2} = \eta_{ely} p_{ely,t}^E \quad \text{Eqn 3.5.1.}$$

where $\eta_{ely} = 0.70$ represents the assumed lower-heating-value efficiency of PEM electrolysis by 2035, which is consistent with performance projections due to improvements in membrane durability and stack design.

Electrolysers operation is also further constrained by its installed capacity K_{ely}

$$0 \leq p_{ely,t}^{H_2} \leq K_{ely} \quad \text{Eqn 3.5.2.}$$

Electrolysers are assumed to operate without a minimum load requirement or ramp-rate limitations. This allows the model to adjust electrolyser utilisation in response to the availability of low-cost renewable electricity.

The capital expenditure for PEM electrolysers is set at

$$c_{ely}^{CAP} = 700 \text{ EURO/kW} \quad \text{Eqn 3.5.3.}$$

which falls within the cost forecast range for large-scale PEM deployment in the mid-2030s. [22]

3.6 Hydrogen Demand Adjustments

Hydrogen demand within this model is specified exogenously at country level based on each country's national hydrogen roadmaps. For each country, a total annual hydrogen requirement E_{H_2} in (MWh_{H_2}) is imposed to ensure the amount of hydrogen that has to be produced and utilised in the power sector by 2035. Rather than assuming that the demand for hydrogen is constant over time, a temporal profile is shaped according to each country's hydropower availability, shown in Figure 3. To construct this profile, an aggregated hydropower output from all hydro generators within a country is first normalized into a hydro availability index, $h(t)$, which varies between 0 and 1. It will then be converted into a weighing function shown in Equation 3.4.1.

$$w(t) = \frac{h(t)}{\sum_t h(t)} \quad \text{Eqn 3.4.1.}$$

This weighing function index distributes the weight proportionally across all snapshots while maintaining the seasonal shape of hydro generation. Months with higher hydro potential will receive higher weights and vice versa. Furthermore, the total annual hydrogen requirement is allocated across the year according to Equation 3.4.2.

$$P_{H_2}(t) = E_{H_2} w(t), \quad \text{Eqn 3.4.2.}$$

Within PyPSA, this profile is implemented as a must serve hydrogen load on each country's hydrogen bus. As hydrogen storage is not included within our model, electrolysers would have to produce hydrogen within each snapshot in quantities that exactly match the hydrogen demand $P_{H_2}(t) = E_{H_2}$, while fuel cells then reconver hydrogen to electricity in the same period. This ensures that there is a strict temporal coupling between hydrogen production and use.

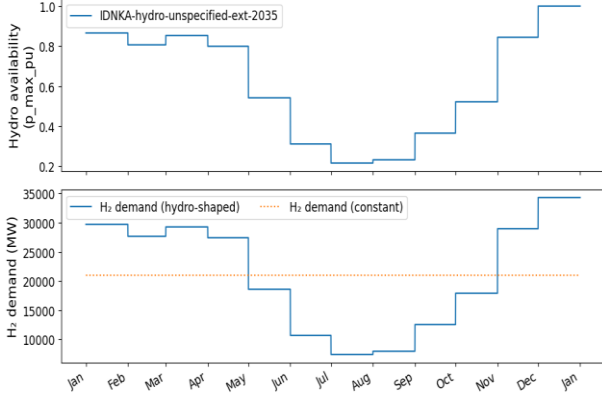


Figure 3: Hydropower availability and H₂ demand in IDN (2035).

3.7 Fossil Capacity and Renewable Expansion

To ensure that the model shows a realistic medium-term decarbonization trajectory, fossil-fuelled generation capacity is treated as fixed and non-expandable. Thus, installed capacity of coal, gas and oil units are imposed as a non-expandable bound within the new optimisation. Specifically, the installed capacity K_i of each fossil unit $i \in F$ is constrained to its upper bound equal to the existing installed capacity K_i^{ref} shown in Equation 3.5.1.

$$K_i \leq K_i^{ref}, \quad \forall i \in F \quad \text{Eqn 3.5.1.}$$

This constraint ensures that the units can still operate and contribute to the system, but it is prohibited from investing in any additional fossil capacity. Preventing unrealistic responses to hydrogen-driven electricity demand, such as expanding coal or gas infrastructure, which would contradict to national energy transition strategies and commitments.

VRE technologies, they are allowed to expand endogenously within the optimisation. Renewable capacity for each technology $j \in R$ is represented as the sum of existing capacity and new-build potential shown in Equation 3.5.2

$$K_j = K_j^{ex} + K_j^{new}, \quad \forall j \in R \quad \text{Eqn 3.5.2.}$$

where new capacity satisfies:

$$0 \leq K_j^{new} \leq \bar{K}_j^{pot}, \quad \forall i \in R \quad \text{Eqn 3.5.3.}$$

$\bar{K}_j^{pot}$ represents the technology-specific upper bound that may reflect technical resource potential or explicit policy making limits. This structure allows the model to optimise renewable capacity where cost-effective, thus ensuring the model is dominated by low-carbon technologies in 2035.

3.8 Optimisation Framework

This optimisation is formulated as a linear least-cost generation expansion and dispatch model using PyPSA, solved over a year with a 2-hour temporal resolution. The objective of this model is to identify the least-cost

combination of electricity generation, hydrogen production and transmission flows required to meet both the electricity generation and hydrogen demand within ASEAN. This is minimised via simultaneous optimisation of investments and operational decisions as expressed in Equation 3.6.1.

$$\min_{[K_g^{new}, p_{g,t}, F_{l,t}]} \left[\sum_g C_g^{Cap} K_g^{New} + \sum_{g,t} C_g^{var} p_{g,t} + \sum_{f,t} C_f^{fuel} f_{f,t} \right] \quad \text{Eqn 3.6.1.}$$

where K_g^{New} denotes the investment in new capacity of technology g , $p_{g,t}$ is the dispatch from generator g at time t , and $f_{f,t}$ is the fossil fuel consumed by unit f . The coefficients C_g^{Cap} , C_g^{var} , and C_f^{fuel} represents the annualized capital cost, variable cost and fuel cost respectively.

Furthermore, operational feasibility is ensured by a series of system-wide constraints. Firstly, generator dispatch is bounded by the installed capacity and availability,

$$0 \leq p_{g,t} \leq K_g \cdot a_{g,t} \quad \text{Eqn 3.6.2.}$$

Where K_g represents the installed capacity of generator g , and, $a_{g,t}$ is the active power availability per unit of the nominal power at snapshot, t . Fossil-fuelled generators K_g are fixed at 2023 capacities, while VRE may expand through optimisation. Electricity supply must meet its traditional demand and also additional demand for hydrogen production. Hence, this can be enforced with a nodal power balance constraint applied at each bus and snapshot:

$$\sum_{g \in \mathcal{G}_n} p_{g,t} + \sum_{\ell \in \mathcal{L}_n} F_{\ell,t} = d_{n,t} + P_{H_2,n}(t) \quad \text{Eqn 3.6.3.}$$

where $d_{n,t}$ is the electrical load at node n . $P_{H_2,n}(t)$ is the electricity used by electrolyzers for hydrogen production at the node, while $F_{\ell,t}$ shows the power flow on each connected transmission lines.

Transmission flows are limited its thermal capacity constraints. For each transmission lines l , the power flow has to satisfy Equation 3.6.4.

$$-\bar{F}_l \leq F_{l,t} \leq \bar{F}_l \quad \text{Eqn 3.6.4.}$$

where $\bar{F}_l$ is the maximum allowable flow. The model uses the BAU transmission configuration from TZ-APG, with all lines fixed to their existing circuit, thus no new transmission expansion occurs.

4. Results

4.1 System-Level & Country-Level Impact of Hydrogen Integration

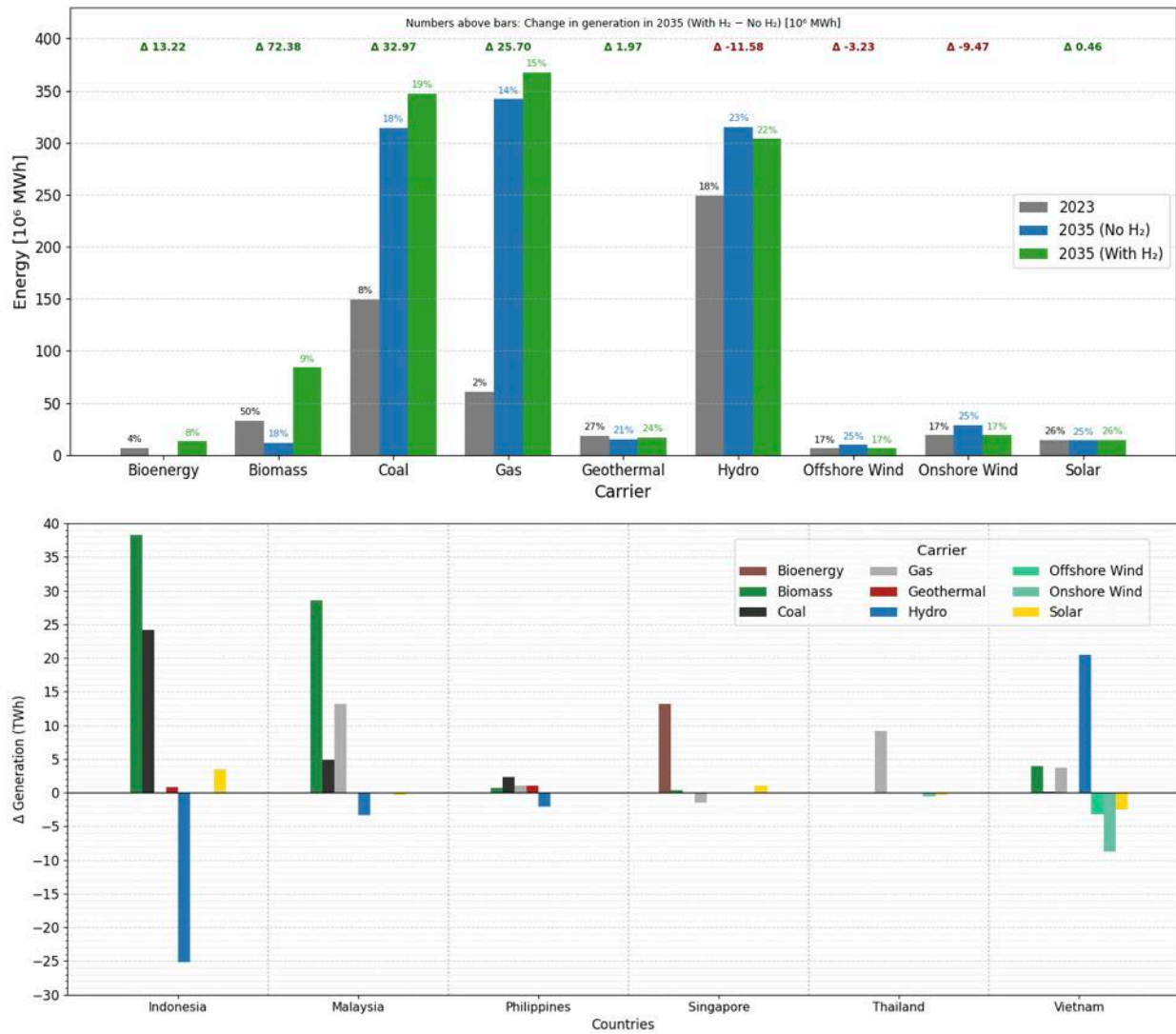


Figure 4: Generated energy by carrier in 2023 vs 2035

Figure 5: Country breakdown of change in generation

The introduction of hydrogen production in 2035 has produced a significant shift in ASEAN's generation mix, as shown in *Figure 4* and *Figure 5*. At a system-level, the aggregate generation increases across nearly all carrier types, with a significant rise of 72.38 TWh in biomass, 32.97 TWh in coal, and 25.70 TWh in gas. On the other hand, hydropower decreased significantly at 11.58 TWh while offshore and onshore wind both declined by 3.23 TWh and 9.47 TWh respectively.

This system-wide pattern is reinforced by the country-level response. Countries like Indonesia, Malaysia and the Philippines have seen a steep reduction in hydropower with -25.00 TWh, -4.00 TWh and -2.00 TWh respectively. In contrast, biomass has increased sharply to offset this decline with an increase of 38.00 TWh in Indonesia, 28.00 TWh in Malaysia and 1.00 TWh in the Philippines. Coal and gas increased modestly within these three countries to provide additional balancing flexibility within the constraints of fixed fossil capacity.

Furthermore, limited renewable resource access countries like Singapore almost exclusively increase generation via bioenergy by about 13.00 TWh in order to meet the hydrogen demand, with minimal change to other carriers. In addition, Thailand has seen a dominant increase in gas with about 9.00 TWh, while variable renewables remain largely unchanged. Vietnam diverges from the regional trend as the only country where hydropower has seen a significant increase, of about 21.00 TWh. Even so, biomass and coal had to be increase by 4.00 TWh and 3.50 TWh in order to fulfil the entire hourly hydrogen load.

4.2 Cost Impact from Hydrogen Integration

The integration of hydrogen production into ASEAN power system has led to a substantial increase in total system cost in 2035. As shown in Table 2, the total annual system cost rose from €111.75 billion per year in the No-H₂ scenario to €145.41 billion per year for the hydrogen integrated scenario, an increase of about €33.66 billion per year. This change in cost reflects on the structural transformation induced by the large-scale deployment of

electrolysers and additional electricity required to supply hydrogen production.

Table 2: Total system cost with and without hydrogen integration in €/b/yr

Scenario	CAPEX	OPEX	Total
No H ₂ 2035	95.20	16.55	111.75
With H ₂ 2035	118.18	27.23	145.41
Change	+22.98	+10.68	+33.66

4.2.1 Total System Generation

Annualised CAPEX has increased from €95.25 billion per year to €118.18 billion per year, indicating an increase of 68%. Table 3 shows the breakdown of an unevenly distributed CAPEX drivers, which is consistent with the model's design of fixing fossil and transmission capacities while the renewable generation and hydrogen infrastructure were extendable.

Table 3: Annualised CAPEX by technology in 2035 in €/b/yr

Group	No H ₂	With H ₂	Δ CAPEX
Fossil generators	41.79	41.79	0.00
Renewable generators	48.70	56.29	+7.59
Hydrogen chain (links)	0.00	15.39	+15.39
Network (links)	4.71	4.71	0.00
Total CAPEX	95.20	118.18	+22.98

The largest component of the CAPEX increment was the hydrogen chains, which accounted for €15.39 billion/year, which was approximately 46% of the total increase in system cost. This includes costs such as the installation of new electrolysers, and all associated linking infrastructure required for electricity-to-hydrogen conversion. As hydrogen demand is exogenously imposed, the system has to install sufficient electrolyser capacity to meet this requirement, thus making the hydrogen chain CAPEX proportional to the scale of hydrogen production.

The second major contribution is the additional investment required for renewable generation, which is approximately €7.59 billion/year. This expansion is structurally necessary because renewables such as biomass and bioenergy have increased drastically in countries like Singapore, Indonesia and Malaysia. On the other hand, the CAPEX for fossil generation and transmission network remains unchanged across all scenarios as the capacity for coal, gas and oil is held at constant accordance to the model's assumption of non-extendability, hence remaining at €4.71 billion/year.

Table 4: Breakdown of OPEX with and without hydrogen integration in €/b/yr

Group	No H ₂	With H ₂	Δ OPEX
-------	-------------------	---------------------	--------

Fossil generators	16.30	17.78	+1.47
Renewable generators	0.25	0.56	+0.31
Hydrogen chain (links)	0.00	8.89	+8.89
Network (links)	0.00	0.00	0.00
Total OPEX	16.55	27.23	+10.68

As seen in Table 4, the pattern observed in OPEX is very similar to CAPEX, where OPEX grows from €16.55 to €27.23 billion per year, driven primarily by the operating cost for hydrogen chains, reflecting the continuous electrolyser operation and associated conversion processes. The notable difference is the rise in OPEX for fossil generators at €1.47 billion per year. Although no fossil fuel units are being built, the higher utilization of the existing fleet resulted in an increase in fuel burn and, therefore, higher operating costs.

5. Discussion

Overall, the levelized cost of hydrogen (LCOH) for the 2035 hydrogen scenario is around €4.2/kg, which corresponds to around USD \$5/kg at current exchange rates. This is significantly higher than the USD \$ 2.0-2.5/kg range that ERIA has highlighted as being competitive for renewable hydrogen globally. [8] This suggests that the results are interpreted as a conservative upper bound, as all additional system costs are effectively attributed to hydrogen production. Hence, under our current assumptions, hydrogen production and integration within ASEAN remain relatively expensive, and bridging this gap would require lower electrolyser CAPEX, cheaper and more available renewable electricity, and coordinated regional planning for shared infrastructure.

Furthermore, the integration of hydrogen production into ASEAN's power grid has led to a reallocation of generation away from hydropower toward more dispatchable sources. In particular, total hydropower output fell significantly under the hydrogen scenario, while generation of bioenergy rose significantly to compensate. Coal- and gas-fired generation also increased modestly, indicating that additional capacity required for the electrolysers was obtained from spare capacity in existing fossil fuel plants. Even variable renewables such as wind saw a slight decline in output, indicating that without storage, the model could not fully leverage intermittent resources for hydrogen production.

For country-specific results, the model has been shown to amplify the existing structural features of the national power systems rather than making them more similar. Countries with higher hydropower, such as Malaysia, Indonesia, and the Philippines, experienced a significant decline in hydro generation and countervail with biomass-based electricity. In contrast, countries with limited renewable resources or rigid generation profiles

turned to what was available domestically. For example, Singapore, which lacks land for solar and wind, relies almost entirely on bioenergy for hydrogen production, even though this may be difficult to realise in practice given its resource constraints. Thailand uses its existing national gas fleet more heavily. Renewables in those countries mostly remained unchanged due to practical limits in scaling them quickly under the given constraints. An interesting outlier was Vietnam, which, unlike its neighbours, was able to increase hydropower output to supply hydrogen, yet still had to bolster some generation via biomass and coal to fully cover its new hydrogen load. Hence, the model was able to allocate dispatchable renewables to nations that are rich in renewables, whereas others may fall back on fossil fuels if alternatives are scarce.

Furthermore, it is important to note that fossil-fuel capacities were set as non-extendable, but the capacity of the constraint itself never became binding in either scenario. This is because the existing fossil capacity still had spare headroom, allowing a rise in fossil output from current plants, rather than building new ones. Renewables were modelled with no fuel cost compared to fossil fuels; thus, the model tends to ramp up biomass-based generation first, using it as the primary dispatchable option.

The absence of hydrogen storage in the current model is another key simplification that tends to bias the system towards dispatchable carriers. Hydrogen storage enables electrolyzers to operate when electricity is cheap or in surplus, separating hydrogen production from its use. Studies have shown that combining electrolyzers with hydrogen storage reduces curtailment and lowers balancing requirements from thermal plants. [23] However, in this project, hydrogen was produced and used within the same two-hour timestep, eliminating the potential storage role. This was a deliberate modelling choice for clarity, but it also means the result should be interpreted as an upper bound on the cost. In a realistic scenario with buffer storage, we would expect the system to exploit low-cost periods for hydrogen production, thereby reducing reliance on expensive dispatchable or fossil generation.

The model also enforced the BAU topology and did not allow any expansion of the cross-border grid beyond today's level. This has provided a stable reference for analysis, but it has limited the model's ability to share resources, especially those that are curtailed. Each country had to balance its own hydrogen supply internally, given the constrained grid. This fixed transmission network has most likely led countries like Singapore to heavily depend on domestic bioenergy and Thailand to rely on its own gas. This may therefore overemphasize local solutions and underplay a regional optimal solution, unable to realize the cost savings that a future expanded transmission network might offer.

6. Conclusion and Outlook

Overall, this report shows that imposing hydrogen targets on ASEAN's power system would substantially reshape the regional generation mix and increase total system costs in 2035. When hydrogen demand is considered, hydropower capacity would decrease, while biomass capacity would increase, due to the additional CAPEX and OPEX required. However, several structural limitations within the present framework have shaped these results. A key limitation is that the model generates hydrogen domestically, with no representation of cross-border pipelines. China's emerging hydrogen plans have shown that once hydrogen flows are large and continuous, dedicated pipelines can minimize levelized transport costs compared to trucked compressed or liquified hydrogen. [24] Techno-economic assessments of industrial clusters have shown that for a moderate demand of 80 GWh, pipelines can undercut on-site electrolysis by 14-24 ct/kWh, compared to 29-49 ct/kWh. [25]

Secondly, the geographical scope of this study is confined to ASEAN and therefore neglects China's role, even though it shapes ASEAN's power sector and hydrogen prospects. China's Global Energy Interconnection (GEI) plans to move renewable power to demand centres worldwide, including the proposed China – Southeast Asia link. [26] Trajectories have shown that around 96% of China's electricity generation comes from renewables by 2060, while ASEAN Centre for Energy estimates that its own transition would require approximately USD 3.7-11.9 trillion in energy investments by 2050. [27] [28] Thus, if such interconnections and co-developments do materialize, the overall cost of meeting hydrogen targets could differ substantially from this report.

This report indicates that large-scale hydrogen deployment in ASEAN cannot be assessed in isolation from regional power trade, midstream infrastructure, and the wider geopolitical context. The result from our model is best viewed as an upper-bound stress test of what it might cost to meet hydrogen targets, largely through domestic generation and the existing grid. Future modelling should take these limitations into account to ensure robust and accurate hydrogen deployment strategies that support rather than strain ASEAN's net-zero transition.

7. Acknowledgements

The authors would like to extend their gratitude to Sridharan, Vignesh, and Schirren, Lukas, for their continuous support and guidance throughout this research project.

References

- [1] ASEAN Centre for Energy (ACE) (2025) *ASEAN Low-Carbon Energy Technologies Roadmap (ALERT) – Phase I: ASEAN’s Long-Term Strategy on Hydrogen and Ammonia*. Economic Research Institute for ASEAN and East Asia (ERIA), Jakarta.
- [2] Organisation for Economic Co-operation and Development (OECD) (2023) *The Climate Action Monitor 2023: Providing Information to Monitor Progress Towards Net-zero*. OECD Publishing.
- [3] International Energy Agency (IEA) (2020) *Global Hydrogen Review 2020*. OECD Publishing. Available at: <https://www.iea.org/reports/global-hydrogen-review-2020>
- [4] TransitionZero (2024) *From Vision to Voltage: Open-Access Modelling of the ASEAN Power Grid (TZ-APG v1)*.
- [5] ASEAN Centre for Energy and HAPUA (2021) *ASEAN Interconnection Masterplan Study (AIMS) III Report: Phase 1 – Capacity Expansion Planning; Phase 2 – Grid Performance Analysis*. Endorsed by the 39th ASEAN Ministers on Energy Meeting (AMEM), Brunei Darussalam, 15 September 2021.
- [6] UK Energy Storage (2025) UK Energy Storage: The UK’s largest underground salt cavern hydrogen storage projects – Dorset & East Yorkshire. Available at: UK Energy Storage website
- [7] Gawlick, J. and Hamacher, T. (2023) ‘Impact of coupling the electricity and hydrogen sector in a zero-emission European energy system in 2050’, *Energy Policy*, 180, 113646. doi:10.1016/j.enpol.2023.113646
- [8] Purwanto, A. J. and Rusli, R. D. (eds.) (2023) *Hydrogen Demand and Supply in ASEAN’s Industry Sector: Current Situation and the Potential of a Greener Future*. Economic Research Institute for ASEAN and East Asia (ERIA).
- [9] ASEAN Centre for Energy (ACE) (2024) *8th ASEAN Energy Outlook 2023–2050*. Jakarta: ASEAN Centre for Energy.
- [10] ERIA (2019) ‘Estimation of hydrogen demand potential in the East Asia Summit region’, in Kimura, S. and Li, Y. (eds.) *Demand and Supply Potential of Hydrogen Energy in East Asia*. ERIA Research Project Report FY2018 no. 01. Jakarta: ERIA, pp. 53–139.
- [11] International Energy Agency (IEA) (2024) *Global Hydrogen Review 2024*. Paris: IEA. Available at: <https://www.iea.org/reports/global-hydrogen-review-2024>
- [12] Genovese, M., Schlüter, A., Scionti, E., Piraino, F., Corigliano, O., Fragiaco, P. et al. (2023) ‘Power-to-hydrogen and hydrogen-to-X energy systems for the industry of the future in Europe’, *International Journal of Hydrogen Energy*, 48(44), pp. 16545–16568. doi:10.1016/j.ijhydene.2023.01.194
- [13] Zeyen, E., Victoria, M. and Brown, T. (2023) ‘Endogenous learning for green hydrogen in a sector-coupled energy model for Europe’, *Nature Communications*, 14, 3743. doi:10.1038/s41467-023-39397-2
- [14] Phoumin, H., Kimura, F. and Arima, J. (2021) ‘Potential green hydrogen from curtailed electricity in ASEAN: The scenarios and policy implications’, in Phoumin, H., Taghizadeh-Hesary, F., Kimura, F. and Arima, J. (eds.) *Energy Sustainability and Climate Change in ASEAN*. Singapore: Springer, pp. 195–216. doi:10.1007/978-981-16-2000-3_8
- [15] Tomlinson, A. (n.d.) *Indonesia’s National Hydrogen Strategy: Paving the Way for Green Energy Transition*.
- [16] Global Green Growth Institute (GGGI) (2025) ‘Indonesia strives for an ecosystem of clean hydrogen for a greener future’, 1 June. Available at: <https://ggi.org/indonesia-strives-for-an-ecosystem-of-clean-hydrogen-for-a-greener-future/>
- [17] ChemAnalyst (2024) *Malaysia Hydrogen Market Analysis: Industry Market Size, Plant Capacity, Production, Operating Efficiency, Demand & Supply, End-user Industries, Sales Channel, Regional Demand, Company Share, Foreign Trade, Manufacturing Process, 2015–2034*. January. Available at: <https://www.chemanalyst.com/industry-report/malaysia-hydrogen-market-740>
- [18] Malaysian Investment Development Authority (MIDA) (2024) ‘Malaysia aims to produce 2 million tonnes of hydrogen annually by 2030’, 27 November. Available at: <https://www.mida.gov.my/mida-news/malaysia-aims-to-produce-2-million-tonnes-of-hydrogen-annually-by-2030/>
- [19] International Energy Agency (IEA) (2025) *Global Hydrogen Review 2025*. Paris: IEA. Available at: <https://www.iea.org/reports/global-hydrogen-review-2025>
- [20] Future Market Insights (2025) *Hydrogen Electrolyzer Market: Size and Share Forecast Outlook 2025 to 2035*. Future Market Insights, 5 December.
- [21] Buttler, A. and Spliethoff, H. (2018) ‘Current status of water electrolysis for energy storage, grid balancing and sector coupling via power-to-gas and power-to-liquids: A review’, *Renewable and Sustainable Energy Reviews*, 82(3), pp. 2440–2454.
- [22] Schmidt, O., Gambhir, A., Staffell, I., Hawkes, A., Nelson, J. and Few, S. (2017) ‘Future cost and performance of water electrolysis: An expert elicitation study’, *International Journal of Hydrogen Energy*, 42(52), pp. 30470–30492.
- [23] Ge, P., Hu, Q., Wu, Q., Dou, X., Wu, Z. and Ding, Y. (2020) ‘Increasing operational flexibility of integrated energy systems by introducing power to hydrogen’, *IET Renewable Power Generation*, 14(3), pp. 372–380. doi:10.1049/iet-rpg.2019.0663
- [24] Zhao, S., Wang, W., Li, Y. and Liu, C. (2025) ‘Pipeline transportation strategy and key technological breakthroughs for high-quality development of hydrogen energy industry under the dual carbon goals’, *Journal of Pipeline Science and Engineering*. Advance online publication. doi:10.1016/j.jpse.2025.100391
- [25] Cvetkovska, R., Wechner, L. and Kienberger, T. (2025) ‘Techno-economic assessment of hydrogen supply solutions for industrial site’, *International Journal of Hydrogen Energy*, 104, pp. 611–622.

- [26] Delina, L. L. (2021) ‘Promises and pitfalls of China–Southeast Asia energy connectivity’, *Energy Strategy Reviews*, 33, 100574. doi:10.1016/j.esr.2020.100574
- [27] ASEAN Centre for Energy (2025) ‘ASEAN–China knowledge exchange on low-carbon energy investment’, *ASEAN Centre for Energy*, 28 July. Available at: <https://aseanenergy.org/post/asean-china-knowledge-exchange-on-low-carbon-energy-investment/>
- [28] Energy Research Institute of Chinese Academy of Macroeconomic Research (2024) *China Energy Transformation Outlook 2024 (CETO24)*. China Energy Transformation Program.

Dynamic Fouling Modelling in Skimmed Milk Powder Manufacturing

Aarna Jindal and Koyna Bajaj

Department of Chemical Engineering, Imperial College London, London, United Kingdom.

Abstract: Skimmed milk powder (SMP) plays a crucial role in meeting global nutrition needs due to its long shelf life, but this comes with a high energy and carbon footprint compared to liquid milk. This report aims to quantify the effect of fouling and cleaning-in-place (CIP) on energy use and cost in SMP manufacture by converting a steady-state SuperPro Designer process simulation into a time-variant energy model. The full process (pasteurisation, standardisation, evaporation, spray drying and CIP) was first simulated at steady state under various process efficiencies to cover a broad range of potential operating scenarios. Outputs were then combined with empirical relationships that captured the result of fouling on process efficiency. For pasteurisation, a fouling kinetic predictive model combined with heat-balance relationships was used to relate duty to effective efficiency. The evaporator fouling was represented indirectly using a predictive model that captures the reduction in evaporation percentage due to ineffective heat transfer, which leads to increased dryer steam demand. Embedding these relationships in a single time-dependent framework allowed total energy demand and product-embedded energy to be tracked as fouling progressed and CIP durations were varied. The model indicates that evaporation and drying together account for close to 80% of heating duty and yield a base case of about 23 MJ/kg of milk powder. Increasing operating time by 30% with an aim to increase plant productivity reduces specific energy by 0.9%. Furthermore, a 10-minute reduction in caustic and acid CIP steps lowers energy by 0.28 MJ/kg and operating cost by around \$96,000 per year for a medium-sized plant.

Key Words: Skimmed Milk Powder, Dairy Processing, Fouling, Cleaning-in-place, Dynamic Modelling, Efficiency

1. Introduction

Global dairy demand is undergoing a significant surge, mainly propelled by population growth and the rising health and nutritional awareness. As incomes rise, particularly in populous regions like South Asia, dairy is increasingly valued as essential to a "healthy, balanced and nutritious" diet (OECD-FAO Agricultural Outlook 2025-2034, 2025). Within this context, longer shelf-life products play a critical role: dairy powders have a longer life than liquid milk and require no refrigeration (Tetrapak.com, 2025), hence offering reliable access to nutrition even in regions without steady electricity, refrigeration or safe transport. Thus, the demand for extended shelf-life dairy products (e.g. milk powder, whey protein powder and infant formulas) is increasing rapidly, with the milk powder market alone projected to grow from \$35.11 billion in 2025 to \$42.23 billion by 2030 (www.mordorintelligence.com, n.d.). Demand shifts like such, alongside strict food hygiene and processing protocols have put augmenting pressure on the dairy industry's energy consumption, which is already extremely energy intensive (Malliaroudaki *et al.*, 2022).

A study conducted in the UK found that, second to cheese, powdered milk consumes the most energy in the food sector – 10.3 MJ per kg (Corigliano and Algieri, 2024) and is responsible for 15% of total energy used in the dairy industry (Moejes and van Boxtel, 2017). This has led to substantial environmental impacts, for example, one LCA determined that 17.24 kg CO₂e was

emitted per kg milk powder produced. For context, typically milk production releases 2.5 kg CO₂e per kg (Wright, 2021). Rising concerns for the environment and public health are now incentivising efforts to decarbonise dairy processing and, henceforth, improving the energy efficiency in the dairy industry, specifically milk powder manufacturing.

2. Background

Milk powder production is governed by four sequential steps: pasteurisation, standardisation, evaporation and spray drying (Figure 2.1). Raw milk is pasteurised for microbial safety and then, clarified (removal of solids and debris), cooled, separated into cream and skimmed milk. Post this, the milk is concentrated in a multi-effect evaporator that evaporates 70%–90% of the water, after which the resulting concentrate is atomised in a spray dryer and contacted with hot air to reduce the remaining moisture to 2%–4% and form milk powder. The product is collected, screened for hygiene and product standards, packaged and shipped. Between process stages, cleaning-in-place (CIP) procedures are implemented to remove fouling deposits and maintain hygiene (Uçal *et al.*, 2023). Fouling refers to the accumulation of milk solids (mainly denatured proteins) on heat transfer surfaces, forming an insulation deposit layer that over time reduces the thermal efficiency and increases the pressure drop. Therefore, fouling is a key driver for energy consumption in powder processing.

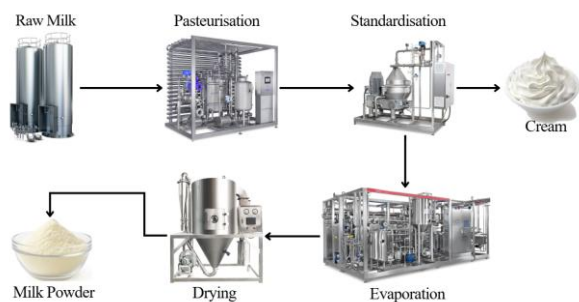


Figure 2.1. Milk Powder Production Process Stage-wise Visual

Apart from fouling, milk powder manufacturing is also energy intensive due to its dehydration steps: evaporation and spray drying. The evaporation and drying of milk to milk powder cumulatively consume 36% of electrical and 91% of thermal energy respectively of the process (Finnegan *et al.*, 2017). An industry report estimates that industrial spray dryers consume on the order of 3–5 MJ per kg of water evaporated (Baker and McKenzie, 2005), and a typical multi-effect milk evaporator uses about 0.5–0.9 MJ per kg of water evaporated (Tsochatzidi *et al.*, 2024). Moreover, there is a trade-off between the stages, where improving evaporation operational efficiency reduces the burden on drying and vice versa. This is because any loss of evaporator performance (e.g. due to fouling) moves extra load onto the dryer, as the dryer now needs to reduce the water content by a higher amount. Since, evaporation and drying dominate the energy footprint, even modest improvements in its energy modelling and efficiency can substantially reduce total plant energy consumption and decarbonisation pressure.

Existing efforts to better the milk powder manufacturing process mostly use process integration and modelling techniques such as pinch analysis and supply chain energy models (SCEM) (Malliaroudaki *et al.*, 2022). These approaches are valuable for mapping overall energy flows and identifying theoretical energy savings, for instance, by optimizing heat exchanger networks in the stages (Kaviani *et al.*, 2022). However, a limitation in these models is that they assume the process operates continuously in steady state. They fail to account for the pertinent issue of fouling and the consequent necessity of periodic cleaning-in-place (CIP) cycles. Fouling drastically reduces heat transfer efficiency over time, thus increasing energy demand as more heat is now required for sufficient heat treatment. Likewise, CIP cycles require significant amounts of energy for heating up the cleaning chemicals and result in longer downtimes. By not integrating fouling and CIP, current models present an incomplete and overly optimistic model of energy consumption, underestimating the total energy intensity and operational costs.

Hence, there is a need for an integrated process model for milk powder manufacturing that accounts for fouling during production, cleaning, and the interdependency of the evaporation and drying processes. This model would serve as a key tool for implementing energy efficiency operation. Therefore, the aim is to convert a steady-state process simulation model to a dynamic mathematical model for quantifying the impact of fouling and cleaning on (the cost and energy of) skimmed milk powder manufacturing. This is especially urgent given the impending climate change impact of the industry and its need to mitigate substantial environmental footprint and move towards net-zero carbon emissions while meeting rising global demand (Malliaroudaki *et al.*, 2022).

3. Methodology

The entire skimmed milk powder production process was modelled using SuperPro Designer (Sánchez *et al.*, 2022), a process simulation software that is used specifically in the analysis and design of food and biochemical industries. SuperPro Designer has an in-built process flow sheet that contains four sequential operations: pasteurisation, standardisation, evaporation and spray drying. It simulates the procedure across all the stages at steady state. The software gives the flexibility to alter each operation discretely for variable input, output and process parameters.

3.1 SuperPro Designer

SuperPro Designer models the production of skimmed milk powder from an input of 100 metric tonnes (MT) of raw milk per batch and produces 9.04 MT of skimmed milk powder per batch. There are 330 batches per year which lead to a production of 2,983 MT of skimmed milk powder per year, representing a medium scale skimmed milk powder plant. Table 3.1.1 summarises the thermal agents required for each section.

Table 3.1.1. Thermal agents per section

Section	Heating Agent	Cooling Agent
Pasteurisation	Hot Water	Chilled Water, CaCl ₂ Brine
Standardisation	Hot Water	
Evaporation	Low Pressure (LP) Steam, Hot Water	Chilled Water
Drying	High Pressure (HP) Steam	

The in-built CIP routines use water, dilute caustic soda (NaOH) and dilute phosphoric acid (H₃PO₄) as the cleaning agents. The simulation has seven CIP schedules, one in pasteurisation (P-4 Mixing), three in

standardisation (P-7 Centrifugation, P-8 Cream Storage, P-9 Mixing), two in evaporation (P-18 Evaporation and P-19 Mixing), and one in drying (P-25 Spray Drying). Each of these CIP schedules run for 2 hours, except the one for P-7, which runs for 1.5 hours. (Appendix Table A.1.1 and A.1.2) mentions further CIP schedule details.

3.2 Fouling Simulation

A bottom-up strategy was used to integrate the dynamic impact of fouling into the overall energy model. First, empirical time-dependent relationships were gathered for each unit operation to quantify how process performance declines as fouling progresses.

In principle, a fully time dependent process model where equipment performance gradually declines as fouling develops would be required to capture the transient behaviour of energy requirements. However, SuperPro Designer does not inherently account for fouling, as it is a steady-state simulation tool and cannot represent dynamic changes in operational efficiency. To address this, fouling effects were introduced by varying the efficiency of key unit operations rather than by modelling deposit growth directly (Appendix A.3).

Consequently, a snapshot-based approach was adopted, in which a series of steady-state simulations at different efficiencies were run. Each simulation represented the plant at a specific point in time and, by extension, at a specific fouling state. For example, a heat-transfer efficiency drop of 1% may correspond to the performance of the pasteuriser after 4 hours of operation. Combining the data from these runs allowed the construction of a time-varying energy model for skimmed milk manufacturing that captured how fouling degrades equipment performance and increases energy consumption throughout an operating cycle.

3.3 Pasteurisation: Energy Modelling

3.3.1 Dynamic Fouling Model

To model fouling in the pasteuriser and its time-dependent effect on heating duty, a dairy fouling predictive model (coupled with heat exchanger energy and mass balance equations) was implemented. Firstly, literature kinetic models were used to predict fouling dynamics during milk heat treatment (Alhuthali *et al.*, 2022). An adjustable fouling-rate scaling parameter in the Biot number-based correlation (Alhuthali *et al.*, 2022) was calibrated to reproduce a performance loss of $\Delta T = 5$ K after 10 hours of operation, which is consistent with industrial standards. A second-order polynomial trendline was fitted to the predicted decline in heat transfer performance (through heat transfer coefficient,

U) over time (t), yielding a compact empirical expression as shown in Eq. 3.3.1.1.

$$\%U_{loss}(t) = -0.0005617t^2 + 0.2475t + 0.0001968 \quad (3.3.1.1)$$

Through Eq. 3.3.1.1, the heating duty demand increase in the pasteuriser was calculated as a function of time. $\%U_{loss}$ can be directly connected to the efficiency of the pasteurisation, η_{pastn} as,

$$\%U_{loss}(t) = 1 - \frac{\eta_{pastn}}{100} \quad (3.3.1.2)$$

3.3.2 Dynamic Energy Model

In SuperPro Designer, the pasteuriser efficiency was varied across the 0.5–1.0 range, and for each value the corresponding heating duty was retrieved, generating a dataset that linked thermal duty to effective efficiency. Regression analysis was then used to obtain an empirical duty-efficiency relationship, which was subsequently combined with the time-dependent fouling model to express pasteurisation energy duty as a function of operating time.

It is to be noted that the chosen efficiency range is wider than typical heat exchanger performance losses but was selected deliberately to generate a broader dataset for regression and analysis.

3.4 Standardisation: Energy Modelling

For standardisation, the energy values were kept constant on an hourly basis. This assumption is justified because the centrifuge (primary unit in standardisation) operates on the principle of gravitational separation rather than heat transfer and hence, it is not significantly affected by fouling. The small amount of heat transfer occurring in other unit operations in this section is insufficient to warrant the modelling of fouling effects. Consequently, standardisation was treated as energetically stable.

3.5 Evaporation: Energy Modelling

In the case of evaporation, the software did not permit direct adjustments to heat transfer efficiency, and modifying the heat transfer area influenced only operating costs rather than the actual energy consumed. As a result, the impact of evaporator fouling could not be modelled explicitly within SuperPro Designer. To address this limitation, evaporator inefficiency was represented indirectly by capturing its downstream effect on the spray dryer. As discussed earlier, there is an inherent trade-off between evaporator and dryer duty: any water not removed in the evaporator must be

removed later in the dryer. Therefore, reduced evaporation efficiency produces a cascading increase in dryer load and energy demand. This indirect approach is valid because it preserves the physical relationship between the two units and reflects how fouling in the evaporator would realistically shift energy requirements across the process.

So, it was assumed that the evaporator has a constant heat supply over time, and the impact of fouling would be realised by inefficiency in the desired water removal (% of water evaporated) over time in the output of the evaporation section.

3.6 Drying: Energy Modelling

3.6.1 Dynamic Fouling Model

The dryer energy duty was dependent on the fouling in the evaporator (modelled by the variable: percent of water evaporated, $Evap\%$) and drying (modelled by the variable: dryer efficiency, η_{dry}). The fouling in evaporator, modelled by $Evap\%$ was found empirically, using simulation data from a study that accounts for fouling build up in a juice evaporator that produces concentrated juice (Cyklis, 2022). The study simulated the evaporator capacity around 10 tonnes/hour which is in the same range as the one used in the present SuperPro Designer simulation. The empirical equation for $Evap\%$ as a function of time is given in Eq. 3.6.1.1 and was assumed to be the same for the simulated milk evaporator.

$$Evap\% = -0.00127t + 0.8804 \quad (3.6.1.1)$$

Then, the spray dryer fouling was assumed to have led to a 2% decline on dryer efficiency over the 6-hour operating period, corresponding to minor but progressive fouling within a single run. Based on literature values indicating that industrial milk spray dryers typically operate with overall thermal efficiencies of about 50%–70% (Cheng, Zhou and Liu, 2018) an efficiency of 0.64 (Appendix A.2) was chosen for this study. This assumption enabled the development of a simple time (t) dependent relationship (Eq. 3.6.1.2) for dryer efficiency (η_{dry}),

$$\eta_{dry} = -0.0033t + 0.6440 \quad (3.6.1.2)$$

3.6.2 Dynamic Energy Model

$Evap\%$ was varied in SuperPro Designer to quantify its effect on dryer steam duty. Starting from the base case (88% water removal in the evaporator), the evaporation rate was reduced to 78% and 68%, thereby increasing the moisture content of the feed entering the spray dryer and raising its energy requirement. For each evaporation

level, simulations were run at three fixed dryer efficiencies (0.5, 0.8 and 1.0), and the corresponding HP steam duties were extracted. The resulting data was then plotted as dryer steam duty versus $Evap\%$ for each efficiency to obtain regression fits.

For the spray dryer itself, fouling was modelled by varying dryer efficiency from 1.0 down to 0.5. SuperPro Designer simulations were run over the 0.5–1.0 range, and at each efficiency level steam duty was recorded, yielding a dataset that relates dryer energy duty to efficiency under progressive fouling. To incorporate the influence of evaporator fouling, this procedure was repeated at different evaporation percentages (88%, 78%, 68%), so that for each combination of dryer efficiency and evaporation level the corresponding steam duty was obtained.

Together, these simulations produced a dataset linking dryer duty simultaneously to dryer efficiency and upstream evaporation fouling. Then, by combining the duty, efficiency, evaporation relationships with the time-dependent fouling descriptions, the energy consumption of drying section was expressed as transient functions, and the indirect effect of evaporator fouling on dryer energy demand was quantified.

3.7 CIP: Energy Modelling

CIP energy modelling was done for all the equipment that had in-built cleaning scheduling. To simplify the calculations for CIP energy, it was assumed that the pump power and energy required for effluent treatment were negligible and the only energy contribution for CIP comes from heating up the cleaning chemicals. Eq. 3.7.1 was used to calculate the energy duty (Q_{CIP}) for each CIP process:

$$Q_{CIP} = \sum_i m C_p \Delta T \quad (3.7.1)$$

where m = mass of the cleaning material (kg/batch), C_p = average specific heat capacity (kJ/kg K), and ΔT = temperature difference (for this, 25 °C was taken as the initial temperature) and i = cleaning chemicals (water, NaOH 2% and H₃PO₄ 2%).

The mass of cleaning chemicals was extracted from the stream summary report generated by SuperPro Designer. The heat capacities of water, and cleaning chemicals were averaged over their respective temperature ranges and have been summarised in Appendix A.4. In this part of the analysis, the in-built CIP durations were not modified, and the operation times were taken as it is; this was defined as the base case.

3.8 Total Energy Modelling

The dynamic energy equations and the values for all the sections: pasteurisation, standardisation, evaporation, drying, and CIP formed a dynamic fouling model that provides the energy consumption for each equipment as a function of time. These equations were used to quantify the effect of longer operating time on the energy demands of the process.

This was done by calculating the total energy required when the operating times for all the sections (except CIP) were increased in 1% intervals till 30%. This implied that the productivity of each batch had also increased by 30%. For example, in the base scenario, 100 MT of milk is pasteurised over 8 hours. When the operating time is increased to 10.4 hours, the system can process 130 MT per batch. This was done for standardisation, evaporation, and drying as well. The CIP time was not increased proportionally when adding its contribution to total energy as it was concluded from the referenced data that the current CIP schedules are sufficient to maintain hygiene standards even at 30% increased flow and operation time (BouMatic, 2021; (Tapflo Pumps UK, 2025).

Moreover, as we are not scaling up the plant, the equipment size is fixed for all unit operations. So, the number of operating days in a year could be increased to accommodate for the longer batch hours (since we increased the operating time by 30%).

For both the pasteurisation and drying processes, the equations provide the power at specific points in time rather than the total energy consumed up to that point. To obtain the cumulative energy usage, the instantaneous values needed to be converted into interval-based energy contributions. Since the model uses time steps of 0.08 hours rather than 1 hour, each value had to be weighted by the corresponding time interval. After applying these interval weights, the cumulative energy was calculated by summing up the energy contributions across all time steps.

The total energy requirements of the milk powder production process were compared against the final product yield to assess how increases in production influence overall energy demand. This was done by calculating the total energy required per unit mass of milk powder produced, and this was defined as the product-embedded energy.

The streams report generated from SuperPro Designer stated that 100 MT of raw milk produces 9.04 MT of skimmed milk powder and this ratio (product

yield) was kept constant when evaluating the quantity of the product at different quantities of raw milk.

3.9 Cost and Energy Cutting Scenarios

Among the five processes considered, CIP is the only procedure that offers a clear opportunity for direct cost reduction, as the referenced data suggest that the software's default schedule results in over-cleaning. It was observed that typical industrial CIP procedures are shorter than the durations assumed in the base scenario of the software model.

So, to evaluate the potential impact of CIP time reduction on energy duty and operational costs, three scenarios were tested. In these, the time allocated to each step involving a 2% NaOH and 2% H₃PO₄ rinse was reduced by 5 minutes (Case 2), and then by 10 minutes (Case 3). The cases are summarised in Table 3.9.1.

Table 3.9.1. *Three CIP scenarios and time change*

CIP step	Time reduction (minutes)		
	Case 1 (base case)	Case 2	Case 3
Rinse 1 (water)	0	0	0
Caustic wash	0	5	10
Rinse 2 (water)	0	0	0
Acid wash	0	5	10
Rinse 3 (water)	0	0	0

These shortened CIP cycles were simulated and the resulting changes in operating costs were extracted from the software. Also, the CIP energy consumption and the product-embedded energy were calculated for each case.

Evaluating energy requirements in this structured way is crucial, as it quantifies how product-embedded energy evolves over the operating cycle and directly informs efficiency-improvement strategies. Incorporating CIP-time reduction scenarios further strengthened the assessment by revealing how shorter cleaning cycles influence CIP-specific energy use, the overall plant energy demand, as well as the operating costs. Together, these insights provide a robust foundation for targeting interventions that could reduce energy consumption, and hence lower operational expenses and enhance the overall sustainability of skimmed milk production.

4. Results and Discussion

4.1 Pasteurisation: Energy Modelling

4.1.1 Pasteurisation Duty against Efficiency

The relationship between thermal duty (kWh/h) against efficiency for pasteurisation was modelled in Figure

4.1.1.1. As efficiency decreased from 1.0 to 0.5, the heating and cooling duties increased.

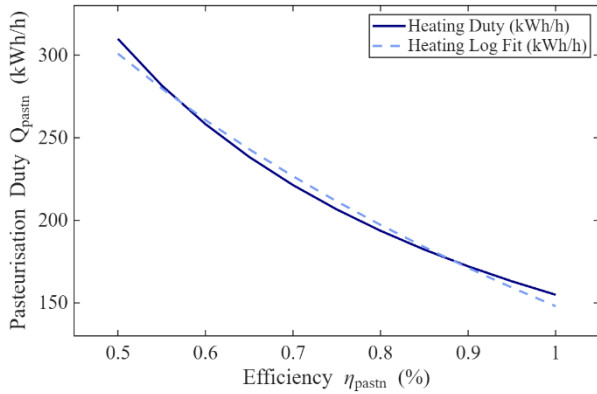


Figure 4.1.1.1. Pasteurisation duty v/s efficiency graph

Over a drop from the 100% to 50% range, the heating duty increased from 155 to 310 kWh/h and the cooling duty from 363 to 440 kWh/h. This reflected the additional energy required to achieve the same thermal treatment when the heat transfer performance is degraded by fouling. A logarithmic regression model was fitted over this data set, showing a strong correlation ($R^2 = 0.99$). This gave empirical equations of the form,

$$Q_{pastn} = -220.6 \ln(\eta_{pastn}) + 148.01 \quad (4.1.1.1)$$

such that, Q_{pastn} is pasteurisation duty, η_{pastn} is pasteurisation thermal efficiency.

4.1.2 Pasteurisation Duty against Time

The empirical model, as stated in Section 3.3.1, was substituted back into the model derived in Section 4.1.1. for pasteurisation duty against efficiency to give a relationship (Eq. 4.1.2.1 and Figure 4.1.2.1) between pasteurisation efficiency and time.

$$Q_{pastn} = -220.6 \ln \left(1 - \left(\frac{-0.0005617t^2 + 0.2475t + 0.0001968}{100} \right) \right) + 148.01 \quad (4.1.2.1)$$

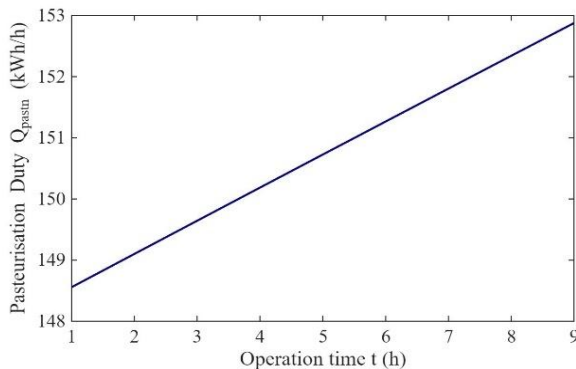


Figure 4.1.2.1. Pasteurisation duty v/s operation time graph

4.2 Standardisation: Energy Modelling

From the reports generated by SuperPro Designer, the energy value extracted was 3368 kWh/h.

4.3 Evaporation: Energy Modelling

From the reports generated by SuperPro Designer, the energy value extracted was 32,570 kWh/h.

4.4 Dryer: Energy Modelling

4.4.1 Dryer Heat Duty against Efficiency and Evaporation %

The influence of dryer efficiency on dryer duty at three *Evap* % levels (68, 78, 88) are observed in Figure 4.4.1.1. There is a general decreasing trend between duty as efficiency increases from 0.5 to 1.0., with the highest gradient of steam duty decline at lowest evaporation % – 68%. There is a strong ($R^2 = 0.99$) logarithmic fit between the efficiency and dryer steam duty.

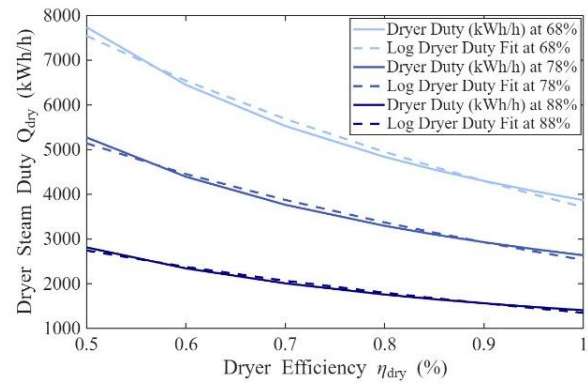


Figure 4.4.1.1. Dryer steam duty v/s dryer efficiency for *Evap* %: 68, 78, 88 graph

Similarly, Figure 4.4.1.2 shows that for three fixed dryer efficiencies (0.5, 0.8, 1.0), dryer steam duty decreases linearly for an increase in *Evap* % from 68% to 88%. A linear model fits for each of the efficiencies ($R^2 = 1$). This justifies using varying *Evap* % to understand the downstream impact that evaporation fouling has on the drying energy requirement.

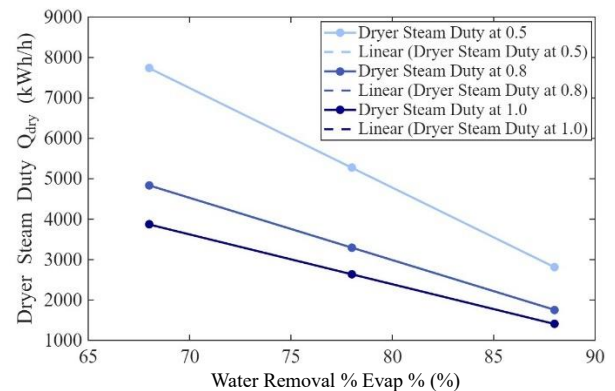


Figure 4.4.1.2. Dryer steam duty v/s water evaporation % for drying efficiencies: 0.5, 0.8, 1.0. graph

4.4.2 Dryer Heat Duty against Time

Therefore, combining the logarithmic and the linear model (Section 4.4.1) a bivariate model for dryer steam duty (Q_{dry}) can be formed as follows:

$$Q_{dry} = c \ln(\eta_{dry}) + d(Evap \%) + e \quad (4.4.2.1)$$

The final model is Eq. 4.4.2.1 for,

$$c = -3767, d = -16451, e = 15360$$

(Further details on the bivariate model reliability can be referred to in Appendix A.6)

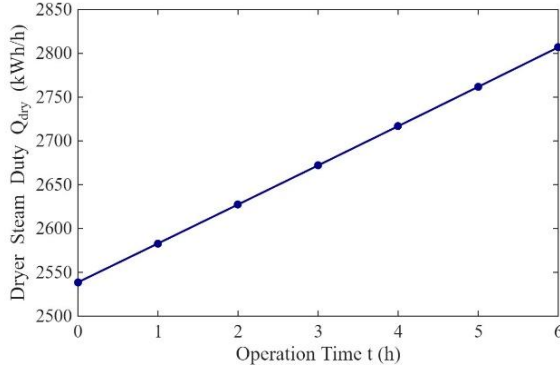


Figure 4.4.2.1. Dryer steam duty v/s time graph

4.5 CIP: Energy Modelling

Table 4.5.1. represents the total calculated CIP energy values for all 7 procedures at the base times (Appendix A.1). It can be observed that P-4 (pasteurisation), P-9 (standardisation), and P-18 (evaporation) account for the largest proportion of the total CIP energy demand, whereas P-25 (drying) is the lowest. This is primarily because the drying branch contains fewer unit operations, and the equipment sizes are significantly smaller compared to those in the upstream liquid-handling sections. Consequently, the corresponding cleaning cycle volumes, flow rates, and heating requirements are lower, resulting in reduced CIP energy consumption for P-25.

Table 4.5.1. Equipment v/s CIP energy

Equipment	CIP Energy Q_{CIP} (kWh/batch)
P-4	757.1
P-7	0.000
P-8	363.8
P-9	728.7
P-18	915.6
P-19	427.4
P-25	108.5
Total	3301

4.6 Total Energy Duty Modelling

The required energy values for each step of the production process were calculated from Eq. 4.1.2, 4.4.2.1, 3.7.1 and Sections 4.2 and 4.3.

In an effort to explore the impact of increasing operating duration, with an aim to increase the plant's productivity, the developed model was used to explore

the total energy demand when operating time increases up to 1.3 times. Figure 4.6.1 shows the variation of total duty, Q_{total} (kWh/MT_{product}) of different processes against an increasing operation time. The graph indicates that evaporation and drying together account for close to 80% of the total heating requirement, highlighting these stages as the key drivers of thermal energy consumption.

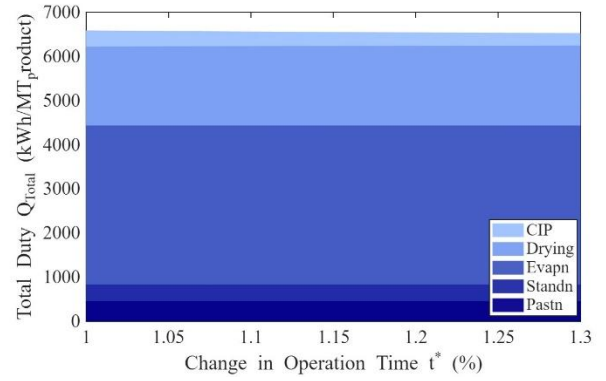


Figure 4.6.1. Heating duty v/s change in operation time graph

Figure 4.6.2 provides a more zoomed-in view of how Q_{total} changes with increasing production time, making the underlying trend easier to observe. The trend shows a decrease in total duty as operating time increases, indicating that longer production runs are more energy efficient.

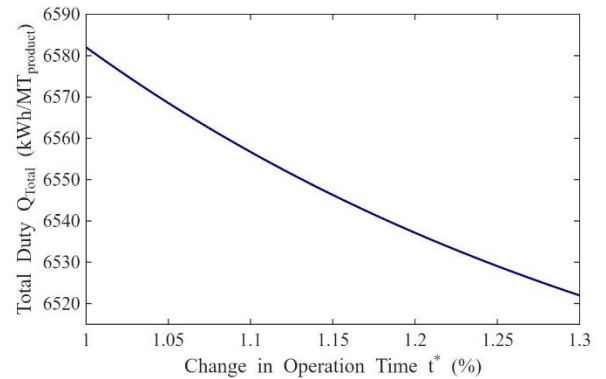


Figure 4.6.2. Total duty v/s change in operation time zoomed-in graph

As the operating time is increased by 30%, the total duty drops by 0.91% from around 6582 kWh/MT_{product} to 6522 kWh/MT_{product}. These energy-saving results can be explained by how CIP energy, Q_{CIP} contributes to the product-embedded energy, Q_{embd} . The energy required for each CIP cycle remains constant, regardless of how much milk is processed. However, when production throughput increases, this fixed CIP energy is spread over more product. As a result, the CIP contribution expressed on a per unit mass basis decreases.

At the same time, fouling does increase the energy consumption during longer operating periods, but within the simulated range, this additional energy

does not yet outweigh the reduction in CIP energy per unit mass. Consequently, the total product-embedded energy continues to decrease with higher throughput.

Although this 0.91% drop might seem like a very small change, it becomes significant as the process is scaled up to commercial standards. For example, in a large-scale plant that produces around 100,000 MT/year (Gea.com, 2025), a 0.91% drop translates to around 21,600,000 MJ on an annual basis.

4.7 Product-Embedded Energy and Cost Analysis

CIP energy duty, Q_{CIP} (kWh/batch) was plotted (Figure 4.7.1) for the three simulated cases corresponding to different levels of CIP time reduction. Each stacked bar is the total Q_{CIP} for a batch, disaggregated by CIP for different equipment (P-4, P-7, P-8, P-9, P-18, P-19, and P-25).

The results demonstrate a decline in total CIP energy demand as cleaning times are shortened. This reduction is primarily driven by the steps with inherently high thermal loads (notably P-4, P-9, and P-18), which contribute substantially to the overall CIP duty. As rinse times are decreased, the energy required for heating and circulating cleaning solutions is proportionally reduced, resulting in lower batch-level energy consumption.

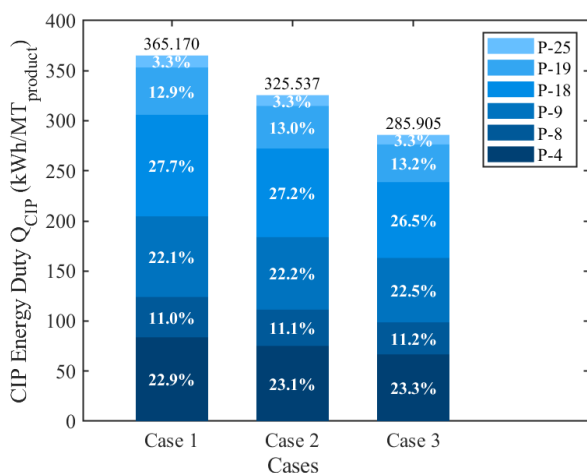


Figure 4.7.1. CIP energy duty equipment shares v/s different cases for CIP time reduction graph

From Table 4.7.1, it can be interpreted that reducing CIP times leads directly to a lower energy requirement per unit of product, Q_{embd} . The shift from Case 1 to Case 2 results in a reduction by 0.14 MJ per kilogram of milk powder, demonstrating the sensitivity of total energy demand to CIP duration. When scaled to an industrial plant, even a one-minute reduction in CIP time across the full process translates to a saving of roughly 200,000 MJ per year. This shows the significant operational and sustainability benefits associated with optimising CIP durations.

Table 4.7.1. CIP scenarios v/s embedded energy v/s operating cost

CIP scenario	Embedded Energy (Q_{embd}) (MJ/kg SMP)	Total annual operating cost (\$*)
Case 1	23.69	30,102,000
Case 2	23.55	30,054,000
Case 3	23.41	30,006,000

*Cost calculations performed for 2022 adjusted \$

It must be noted that although literature reports specific energy consumptions of around 10 MJ/kg, for the factory stages of efficient milk-powder plants, typically considering only evaporation and spray drying with heat integration. In this work, a more complete boundary is used, including pasteurisation, standardisation, evaporation, drying and all CIP duties, without optimised heat recycling. Under these assumptions, the model yields a value around 23 MJ/kg skimmed milk powder, which is higher than benchmark values but consistent with a conservative, non-integrated process.

Moreover, reducing the duration of each caustic and acid wash step by 5 minutes results in an estimated annual cost saving of approximately \$48,000. While this reduction is relatively modest for a facility operating at a multi-million-dollar scale, the benefits of shorter CIP cycles extend beyond the direct operating cost savings. As shown earlier, reducing CIP duration lowers the product-embedded energy, which means the process requires less energy overall and therefore incurs lower energy expenditure. In addition, shorter cleaning cycles reduce the total batch time, allowing the plant to run more batches within the same operating window. Increasing the number of batches raises the annual production volume and ultimately the revenue generated. Thus, although the immediate cost reduction may appear small in isolation, the combined improvements in energy efficiency, operating cost, throughput, and income make CIP time optimisation a meaningful and economically significant enhancement to the process.

5. Conclusions

Therefore, this work developed an integrated SuperPro Designer model of skimmed milk powder production, covering pasteurisation, standardisation, evaporation, spray drying and all associated CIP operations for a 100 MT/batch raw milk input, yielding 9.04 MT/batch skimmed milk powder. An integrated modelling framework was established by combining outputs from simulations with dynamic fouling models, thermal balance equations, and cascading inter-unit effects. Fouling was incorporated explicitly in pasteurisation

and drying via time-dependent efficiency losses, and indirectly in evaporation through reductions in water-removal efficiency that subsequently increased the thermal load on the dryer.

By embedding all unit operations within a unified, time-dependent structure, this study was able to quantify with high resolution how fouling progression and alternative cleaning strategies shape the total energy embedded in the final product. This integrated approach is a key strength of the work, as it captures interactions and cascading effects between unit operations that traditional steady-state modelling cannot represent. The outcome is a flexible, easy-to-use dynamic energy model that can be applied directly without running additional SuperPro Designer simulations, (for 100–130 MT batch size). This significantly reduces computational effort and lowers the entry barrier for users who may not have access to, or familiarity with, process simulation tools. For applications beyond this range of batch size, the methodology provides a transparent, reproducible, and scalable framework that users can readily adapt, making the approach broadly applicable across dairy processing and other manufacturing systems where fouling and energy efficiency are critical.

This modelling approach offers practical value for decision-makers in the dairy industry and beyond. It enables users to assess the impact of fouling, evaluate operational scenarios, and identify opportunities for reducing costs and improving energy efficiency. The same workflow can be applied to other sectors requiring rapid, transparent, and quantitative analysis of profitability, cost structure, and energy demand. While the energy and operating cost savings from CIP-time reduction were modest in this case, the analysis demonstrates that such opportunities are still worth exploring, particularly because benefits scale with production volume and overall plant throughput.

6. Future Scope

Further improvements to the existing approach can be made by developing a dynamic process model as that could increase the energy duty prediction accuracy. Unlike the current steady-state approach, a dynamic model would capture the loss of heat transfer efficiency during continuous operation. This can be further strengthened by incorporating CFD-based fouling simulations, which provide spatially resolved predictions of deposit growth, heat transfer resistance, and flow pattern changes inside heat exchangers and other equipment. Integrating CFD outputs into the dynamic model would allow fouling progression to be represented more realistically, resulting in more accurate

predictions of thermal inefficiency and more reliable optimisation strategies.

Additionally, each process section should be decomposed into its individual unit operations. Modelling unit operations separately not only allows the specific sources of thermal inefficiency to be identified but also highlights operational bottlenecks that are masked when treating sections as a single block. This level of granularity makes it possible to pinpoint where constraints arise; whether from heat transfer limitations, equipment fouling, restricted flow capacity, or cleaning downtime; and therefore, enables more targeted optimisation strategies. Understanding these bottlenecks is essential for improving throughput, reducing unnecessary energy losses, and prioritising upgrades where they yield the greatest impact.

Experimental testing is also required to validate assumptions and generate missing data. Reduced-time CIP cycles should be trialled in a pilot and industrial-scale facility to confirm that hygiene standards remain acceptable when cleaning duration is decreased. Furthermore, empirical data on spray dryer thermal efficiency decay should be collected under real operating conditions, as this information is currently lacking and is essential for accurately modelling long-term performance.

Finally, the model could be expanded to simulate industrial-scale production, up to 40 times the current throughput. Scaling up the batch size will allow the assessment of how energy requirements, fouling behaviour, and CIP demands change at an industrial scale, ensuring that the conclusions drawn from the simulation remain valid for commercial operation.

7. Acknowledgements

We would like to express our sincere gratitude to our industrial supervisor, Dr. Maria Ioanna Malliaroudaki, for her continuous guidance and support throughout this project.

We also acknowledge the use of AI-assisted tools for grammar refinement. All AI suggestions were carefully reviewed, and we confirm that no statements were adopted without verifying their scientific accuracy and ensuring their suitability for this work.

8. Bibliography

Alhuthali, S., Delaplace, G., Macchietto, S. and Bouvier, L. (2022). Whey protein fouling prediction in plate heat exchanger by combining dynamic modelling, dimensional analysis, and symbolic regression. *Food and Bioprocess Processing*, 134, pp.163–180.

<https://doi.org/10.1016/j.fbp.2022.05.009>

Amaury Pérez Sánchez, Nikiforos Misailidis and Petrides, D. (2022). Skim Milk Powder Production – Process Modeling and Techno-Economic Assessment (TEA) using SuperPro Designer. [online] Intelligen, Inc. Available at: https://www.researchgate.net/publication/365236783_Skim_Milk_Powder_Production_-_Process_Modeling_and_Techno-Economic_Assessment_TEA_using_SuperPro_Designer.

Artemis Tsochatzidi, Arvanitidis, A.L. and Georgiadis, M.C. (2024). Model Based Optimization of Energy Consumption in Milk Evaporators. *Processes*, 12(1), pp.209–209. <https://doi.org/10.3390/pr12010209>

Baker, C.G.J. and McKenzie, K.A. (2005). Energy Consumption of Industrial Spray Dryers. *Drying Technology*, 23(1-2), pp.365–386. <https://doi.org/10.1081/drt-200047665>

BouMatic. (2021). Look Towards Optimizing Your CIP Performance. [online] Available at: https://boumatic.com/us_es/expert-blog/look-towards-optimizing-your-cip-performance/ [Accessed 11 Dec. 2025]

Cheng, F., Zhou, X. and Liu, Y. (2018). Methods for Improvement of the Thermal Efficiency during Spray Drying. *E3S Web of Conferences*, 53, p.01031. <https://doi.org/10.1051/e3sconf/20185301031>

Corigliano, O. and Algieri, A. (2024). A comprehensive investigation on energy consumptions, impacts, and challenges of the food industry. *Energy Conversion and Management*: X, 23, p.100661

Cyklis, P. (2022). Effect of fouling on falling film evaporator performance in industrial conditions of fruit juice concentrate production. *Journal of Food Engineering*, 317, p.110884. <https://doi.org/10.1016/j.jfoodeng.2021.110884>

Egan, E.P., Luff, B.B. and Wakefield, Z.T. (1958). Heat Capacity of Phosphoric Acid Solutions, 15 to 80°. *The Journal of Physical Chemistry*, 62(9), pp.1091–1095. <https://doi.org/10.1021/j150567a017>

Finnegan, W., Goggins, J., Clifford, E. and Zhan, X. (2017). Environmental impacts of milk powder and butter manufactured in the Republic of Ireland. *The Science of the Total Environment*. [online] 579, pp.159–168. <https://doi.org/10.1016/j.scitotenv.2016.10.237>

Gea.com. (2025). GEA mandated to build world's largest integrated dairy facility in Algeria. [online] Available at: <https://www.gea.com/en/news/corporate/2025/algeria-worlds-largest-dairy-facility/> [Accessed 9 Dec. 2025]

Kaviani, A., Aslani, A., Zahedi, R., Ahmadi, H. and Malekli, M.R. (2022). A new approach for energy

optimization in dairy industry. *Cleaner Engineering and Technology*, 8, p.100498.

<https://doi.org/10.1016/j.clet.2022.100498>

Malliaroudaki, M.I., Watson, N.J., Ferrari, R., Nchari, L.N. and Gomes, R.L. (2022). Energy management for a net zero dairy supply chain under climate change. *Trends in Food Science & Technology*, 126.

<https://doi.org/10.1016/j.tifs.2022.01.015>

Moejes, S.N. and van Boxtel, A.J.B. (2017). Energy saving potential of emerging technologies in milk powder production. *Trends in Food Science & Technology*, 60, pp.31–42. <https://doi.org/10.1016/j.tifs.2016.10.023>

National Institute of Standards and Technology (2023). Isobaric Properties for Water. NIST Chemistry Webbook, [online] 69. <https://doi.org/10.18434/T4D303>

OECD-FAO Agricultural Outlook 2025-2034. (2025). OECD-FAO Agricultural Outlook. OECD Publishing. <https://doi.org/10.1787/601276cd-en>

Scribd. (2017). Specific Heat of Caustic. [online] Available at: <https://www.scribd.com/document/340696134/Specific-Heat-of-Caustic>

Tapflo Pumps UK. (2025). Clean In Place (CIP) Solutions For Dairy & Milk Processing - Tapflo UK. [online] Available at:

https://www.tapflopumps.co.uk/clean-in-place-cip-solutions-for-dairy-milk-processing/?creative=733682778445&keyword=cip%20procedure%20in%20dairy%20industry&matchtype=e&network=g&device=c&gad_source=1&gad_campaignid=21912124069&gbraid=0AAAAAC5js6ibX3a4Gy7 [Accessed 11 Dec. 2025]

Tetrapak.com. (2025). MILK AND WHEY POWDERED INGREDIENTS | Dairy Processing Handbook. [online] Available at:

<https://dairyprocessinghandbook.tetrapak.com/chapter/milk-and-whey-powdered-ingredients>

Uçal, E., Yildizhan, H., Ameen, A. and Erbay, Z. (2023). Assessment of Whole Milk Powder Production by a Cumulative Exergy Consumption Approach. *Sustainability*, 15(4), p.3475. <https://doi.org/10.3390/su15043475>

Wright, J. (2021). What is the carbon footprint of UK produced milk? [online] Wynnstay Trade Accounts. Available at: <https://www.wynnstay.co.uk/blog/the-carbon-footprint-of-milk/>

www.mordorintelligence.com. (n.d.). Milk Powder Market | Growth | Trends | Forecast. [online] Available at: <https://www.mordorintelligence.com/industry-reports/milk-powder-market>

A. Appendix

A.1 SuperPro Data Base

Table A.1.1 CIP steps v/s CIP time (2h) v/s temperature required

For 2 hours	CIP Time (min)	Temperature required (°C)
Water	15	50
NaOH 2%	30	85
Water	15	50
H ₃ PO ₄ 2%	30	85
Water	30	50

A.1.2 CIP steps v/s CIP time (1.5h) v/s temperature required

For 1.5 hours	CIP Time (min)	Temperature required (°C)
Water	10	25
NaOH 2%	30	25
Water	20	25
H ₃ PO ₄ 2%	20	25
Water	10	25

A.2 Drying Section Efficiency Distribution

In the drying section, the spray dryer accounts for approximately 72% of the total energy demand. Literature reports that industrial milk spray dryers typically operate with thermal efficiencies between 50% and 70%. Because the simulation can only adjust the efficiency of the dryer to represent the inefficiency of the entire drying section, the lower-bound value of 50% was selected to ensure a conservative estimate of performance.

To translate this into an overall section efficiency, the drying section was normalised to receive 100 units of energy, with the spray dryer responsible for 72 units and the remaining sub-operations accounting for 28 units. If the spray dryer operates at 50% efficiency, only 36 of its 72 units become useful energy.

The rest of the section is assumed to operate at full efficiency, contributing its full 28 units. Therefore, the total useful energy becomes $36 + 28 = 64$ units out of the 100 supplied. Using the standard definition of thermal efficiency, the resulting overall efficiency of the drying section is 64%, or 0.64. This value was therefore selected for the study.

A.3 Equipment against Section Comparison

We assume that the efficiency we set for the pasteuriser reflects fouling across the whole pasteurisation section, and we apply the same approach for all other sections. This assumption is justified because each section typically has one major unit operation that accounts for

most of the energy duty. So, modelling fouling on that dominant piece effectively captures most of the fouling behaviour of the section. Also, all equipment in the same section processes identical feed under similar thermal and flow conditions, so the fouling in that main equipment and the whole section are strongly correlated.

A.4 Heat Capacities Calculations

The heat capacities of the cleaning materials: water (National Institute of Standards and Technology, 2023), 2% NaOH (Scribd, 2017), and 2% H₃PO₄ (Egan, Luff and Wakefield, 1958) are summarised in the Table A.3.

Table A.3. Cleaning materials average specific heat capacity

Cleaning material	T_i (°C)	T_f (°C)	C_p (kJ/kg K)
Water	25	50	4.181
NaOH 2%	25	85	4.092
H ₃ PO ₄ 2%	25	85	4.124

where T_i = initial temperature in °C and T_f = final temperature in °C.

A.5 Sanity checks for the developed equations

After developing the energy-efficiency equations for the pasteuriser, standardiser, evaporator and dryer, a sanity check was carried out to verify their accuracy. Representative but random efficiency values were selected: a pasteuriser efficiency of 0.92, an evaporator efficiency of 81%, and a spray dryer efficiency of 0.77.

These values were entered into SuperPro Designer, and the resulting energy report was compared with the energy predicted by the model. The difference between the two was approximately 0.78%, which indicates very strong agreement given the assumptions and simplifications involved.

This was a substantial improvement from earlier iterations in which the error was around 45%. That large discrepancy prompted a detailed review of all calculations and led to the identification and correction of several issues within the equations. therefore, model therefore reflects a carefully validated formulation with close alignment to the simulation outputs.

A.6 Bivariate Model Reliability

Figure A.6.1, shows the comparison between the regression multivariate model for dryer steam duty

against the SuperPro Designer simulation results for each evaporation percents. Across the graph, there are higher deviations for low efficiencies, and a close to exact prediction for 78%, with an overall coefficient of regression, $R^2 = 0.95$.

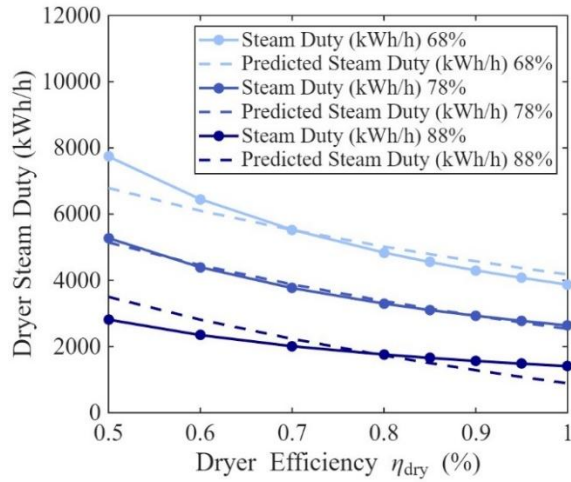


Figure A.6.1. Predicted v/s actual dryer steam duty against Evap % for drying efficiencies: 0.5, 0.8, 1.0.

This confirms that the proposed model reliably represents the relationship between dryer efficiency, evaporation percentage and steam duty with the studied range. Consequently, the regression can be used as a model to estimate dryer energy consumption as fouling progresses, without using the full SuperPro Designer process simulation, and forms the basis for later time and CIP-based analysis.

Assembling Motile Biohybrids via Coordinate Bonding Between DGS-NTA(Ni) Lipids and 6-Histidine

Eylül Akgök

Department of Chemical Engineering, Imperial College London, U.K.

Abstract: Motile biohybrids, combining synthetic giant unilamellar vesicles (GUVs) with genetically engineered *Escherichia coli* (*E. coli*), offer a range of applications particularly in targeted drug delivery. In this study, genetically engineered motile *E. coli* expressing antigen-43 (Ag-43) tagged with 6-histidine (6-His) residues in the N-terminal were attached to DGS-NTA(Ni) functionalised DOPC/cholesterol GUVs via coordinate bonding, a form of biohybrid assembly that remains unexplored. The feasibility of this coordination chemistry was initially tested with 6-His tagged green fluorescent protein (6H-GFP) and were conducted in three different pH conditions (7.5, 8.0, 9.0). 6H-GFP attachment to GUVs showed an optimal signal-to-background fluorescence ratio of 6.682 ± 1.601 at pH 8.0. The vesicle sizes experienced shrinkage, with a narrowed distribution range and decreased average size from $9.22 \pm 4.22 \mu\text{m}$ to $6.22 \pm 1.95 \mu\text{m}$ at ascending external pH. The *E. coli* showed higher swimming speed at pH 7.5 ($17.341 \pm 4.595 \mu\text{m/s}$) compared to pH 8.0 and pH 9.0. The biohybrid assembly demonstrated widespread non-specific interactions complicated by lipid aggregates and electrostatic attraction between negatively charged bacterial membrane and Ni^{2+} , present in the Ni-lipid. These results showed that pH 8.0 is the optimal condition for this biohybrid assembly, balancing *E. coli* speed, Ni^{2+} -His attachment efficiency and vesicle morphology, but it also displays the paramount importance of aggregate-free GUV synthesis.

Keywords: GUVs, electroformation, DGS-NTA(Ni), Ag-6H, GFP, motile *E. coli*, biohybrids

1. Introduction

Giant unilamellar vesicles (GUVs) are adaptable synthetic systems that closely resemble biological cell membranes [1][2]. Due to their functional adaptability and ability to be filled with various cargo [3], they are widely used in research on membrane biophysics [4], biosensors [5], synthetic cell assembly [6], drug delivery [7] and more biotechnological applications [6]. These applications are expanded when pairing the GUVs with motile microorganisms, like bacteria, via encapsulation [8] or membrane attachment to create motile biohybrids. In synthetic biology and microrobotics, motile bacterial biohybrids have become a promising field, developing systems for self-propelling biological applications. The change from passive carriers to micro-swimmers provide new and attractive opportunities for targeted delivery, with another advantage being its biocompatibility [9].

In this paper, the investigation focus was on GUVs formed with DOPC lipids and DGS-NTA(Ni) (Ni-NTA) integrated into its bilipid membrane. These Ni-lipids form coordinate covalent bonds with poly-histidine tagged molecules [10]. This interaction was taken advantage of by modifying a motile *Escherichia coli* (*E. coli*) strain to express a

surface membrane protein, antigen-43 (Ag-43), tagged with 6-histidine (6-His), hence being able to attach itself to cargo loaded liposomes embedded with Ni-lipids. This creates a simple one-step combination between the GUVs and bacteria to form a motile biohybrid, reducing the need of additional lipids to bind the two constructs together. The *E. coli* was also transformed with GFP expressing plasmids, enabling high-contrast fluorescence imaging and clear tracking of the propulsion behaviour.

Initially, 6-His-tagged green fluorescent protein (GFP) was used to check the incorporation of Ni-NTA in the GUV membrane and its binding efficiency in different pH conditions, before the principle was applied to motile bacteria.

2. Background

Bastos-Arrieta et al. [11] has published a comprehensive review, highlighting the advances in bacterial biohybrids and exploring how the characteristics of both bacteria and liposomes can change depending on the surface interactions and external conditions. They further outline how precise control over the attachment of bacteria to synthetic cargo remains a challenge to optimise

propulsion efficiency and functional tunability. Due to the complexity of interaction between the components, various attachment methods have been explored [11].

One assembly method utilising chemical bonding by genetically modifying the bacteria's surface display is leveraged by Schauer et al., where they have engineered *E. coli* to express Ag-43 fused with biotin and thus enabling the bacteria to bind to streptavidin-coated cargo in a controlled manner [12]. This method improves attachment strength, specificity and enables chemotaxis-guided motion, especially as it gives an advantage over other attachment methods that utilise multistep biotin-streptavidin interactions which can decrease the motility of bacteria as their flagellar are highly fragile [12].

Rather than bacteria attachment, Le Nagard et al. [8] has explored motile biohybrid formation by encapsulating bacteria within the liposomes, which led to the lipid vesicles deforming to wrap close enough to the flagellar, producing flagellated-like structures that enabled propulsion.

As different biohybrid construction approaches are explored, Pramanik et al. [10] has systematically characterised the binding chemistry between DGS-NTA(Ni) and 6-His for membrane attachment. Ni-NTA lipids, such as DGS-NTA(Ni) are functionalised lipids that can be integrated into the cholesterol and phospholipid bilayer of synthesised membranes. The nitrilotriacetic acid (NTA) headgroup chelates a Ni^{2+} ion [13] (which itself can form up to 6 bonds with other molecules [14]) through 4 coordinate bonds: three through carboxylate groups, and one through a nitrogen atom [10]. This creates a stable metal coordination complex embedded in the lipid membrane. Proteins tagged with 6-His can bind to the Ni^{2+} ions in the lipid through coordinate bonding [10], filling up the 2 remaining bonding sites on the ion. This occurs through the histidine's imidazole side chain, where its nitrogen donates electrons to the metal ion [10]. Pramanik et al. analysed this system by measuring the binding of His-tagged fluorophores such as GFP (6H-GFP) and FITC (6H-FITC) to DGS-NTA(Ni) functionalised GUVs [10]. Their work explores different GUV synthesis methods, varying the composition of the lipid bilayers, as well as looking at the pH dependence on the attachment and have demonstrated that the maximum binding is exhibited near a pH of 8.5 [10].

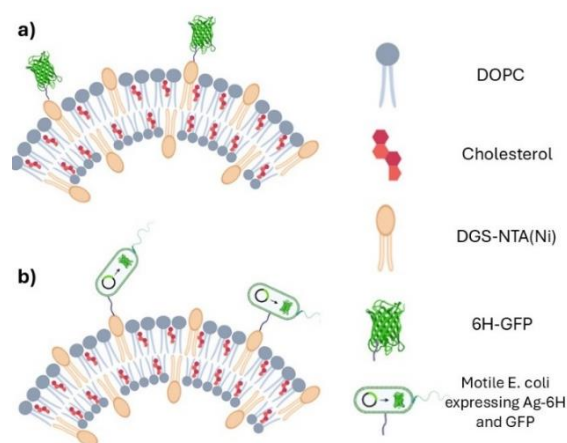


Figure 1. Schematic displaying a) DOPC-based, Ni-NTA integrated bilipid membrane interacting with 6H-GFP b) DOPC-based, Ni-NTA integrated bilipid membrane interacting with Ag-6H expressing *E. coli*

This mechanism provides a logical method for achieving selective bacteria-vesicle attachment in systems where bacteria expresses Ag-43 coupled with a 6-His tag (Ag-6H) on the antigen's N-terminal. This paper combines Ag-6H surface displayed *E. coli* with DGS-NTA(Ni) integrated DOPC GUVs, extending Pramanik et al's [10] protein-level characterisation to motile biohybrids and providing another biohybrid assembly method using chemical bonds alongside the commonly used biotin-streptavidin conjugation [11][12]. The interactions between the Ni-NTA functionalised vesicle surface and 6H-GFP and Ag-6H expressing bacteria are depicted schematically in Figure 1. By exploring pH effects on motility, attachment, and GUV stability, it addresses a gap in distinguishing specific Ni^{2+} -His coupling from aggregation and electrostatic effects in vesicle-bacteria systems.

3. Materials and Methods

3.1. GUV formation via electroformation on ITO glasses

GUVs were prepared via electroformation. The lipid mixture consisted of DOPC (25 mg/mL stock) and cholesterol (25 mg/mL stock) in an 8:2 molar ratio, supplemented with 10 mol% DGS-NTA(Ni) (5 mg/mL stock), and 0.10 mol% of fluorophore Cy5-PE (1 mg/mL stock) for fluorescence visualisation. The final lipid concentration was adjusted to 4mM in a volume of 200 μl with chloroform as the solvent. All lipids were obtained from Avanti Polar Lipids.

The lipid mixture and clean indium tin oxide (ITO) slides were preheated for a few minutes at 75°C using an electric heating plate. A volume of 70

μL of the lipid mixture was pipetted onto one warmed ITO slide and spread into a uniform thin film using a glass coverslip. The film was then further dried with nitrogen gas flow and subsequently kept in a vacuum chamber for 45 minutes to remove the remaining traces of chloroform. The electroformation chamber was assembled by placing the lipid-coated ITO slide against a second clean ITO slide, separated by a cut-out PDMS spacer ($\sim 25 \times 17 \times 2.5 \text{ mm}$) to form a sealed chamber by clipping the slides and the PDMS spacer together. The chamber was injected with 0.5 M sucrose in deionised H_2O .

The slides were connected to a frequency generator (Aim-TTi TG315 3MHz Function Generator), and a sinusoidal alternating electric field with a peak-to-peak voltage of 1.5 V was applied. A frequency of 10 Hz was maintained for 2 hours at 70°C and then the frequency was reduced to 2 Hz, which was applied for 1 hour. GUVs were harvested by carefully separating the slides while tilting the chamber to collect the vesicle-containing sucrose solution into an Eppendorf tube.

It is important to note that it was assumed that the lipid composition in the solution is proportional to the lipid composition of the GUV membrane.

3.2. pH buffer (outer solution) preparation

The outer solutions for the GUV samples consisted of 0.21 M glucose prepared in phosphate-buffered saline (PBS), and this composition was chosen to maintain the osmolarity between the inner and outer solutions. The pH of the outer solution was altered with NaOH and HCl to achieve various buffers of pH 7.5, 8.0 and 9.0. The pH was measured using a pH probe.

3.3. GFP-GUV attachment

Green fluorescent protein (GFP) tagged with 6-His was initially prepared as a 50 mM stock solution. For each experiment, the stock was first diluted to 500 μM in the respective pH buffer/ outer solution (as prepared in section 3.2). This solution was then serially diluted to a final concentration of 5 μM and incubated for 1 hour at room temperature with a 10-fold diluted GUV suspension, using the same pH buffer for dilution. After incubation, the sample was further diluted 20-fold to reduce background fluorescence during imaging.

3.4. Bacteria culturing and biohybrid formation

A motile laboratory prepared *Escherichia coli* MG1655 strain was used. The bacteria were transformed with two plasmids: one expressing GFP, and the other expressing T5- promoter, IPTG-inducible Ag-6H.

Bacteria were streaked onto agar plates and grown overnight to obtain isolated colonies. A single colony was then used to prepare an overnight liquid culture in 10 mL of Luria-Bertani (LB) medium, incubated at 37°C for around 17 hours. The following day, a liquid subculture was prepared in 10 mL of Terrific Broth (TB) medium with 100 μL of the overnight liquid culture, incubated with 10 μL of ampicillin, spectinomycin and Isopropyl β -D-1-thiogalactopyranoside (IPTG) at 30°C for 5-6 hours until it reached an optical density of around 0.5 at a wavelength of 600 nm. This was measured using a BioDrop Spectrophotometer. Ampicillin and spectinomycin were used as selection antibiotics, as the corresponding resistance genes served as selection markers on the plasmids transformed into the bacteria.

500 μL of the cultivated bacteria was centrifuged at 7000 rpm for 5 minutes. The supernatant was discarded, and the cell pellets were gently resuspended in the corresponding outer solution used. The harvested stock GUVs were diluted 100-fold with the bacteria-buffer solution and attachment was monitored with an optical microscope, further detailed in section 3.6.

3.5. Negative test controls

Three negative test controls were focused on to compare the relative attachment of *E. coli* on the GUV membrane: (1) no IPTG was added to bacterial liquid culture, (2) the *E. coli* used in the test did not express 6-His tags, (3) the lipid mixture prepared to synthesise GUVs did not include Ni-lipids, however the mixture still contained a total lipid concentration of 4 mM.

3.6. Microscope analysis

A square PDMS chamber with a volume of approximately 200 μL was cut and filled with the test sample. Prior to use, the chamber surfaces were passivated with a freshly prepared 10% (w/v) bovine serum albumin (BSA) solution in 0.21 M glucose with PBS. The passivation step was carried out for 30–60 minutes to minimise nonspecific adhesion, such as bacterial attachment to the surface.

Passivation also prevents the GUV suspended in the outer solution to interact with the cover slip and prevent bursting. Imaging was performed using a Nikon Eclipse Ti2 optical microscope equipped with 20× and 40× objectives (numerical aperture (NA) values of 0.25 and 0.75 respectively) immersed in air. Excitation was provided by lasers at 475 nm (20% power) and 621 nm (40% power) for GFP and Cy5-PE fluorescence, respectively, with their emission peaks found at 509 nm [15] and 667 nm [16]. For GUV-only analysis, the exposure time was kept at 500 ms for GFP imaging, and 30 ms for Cy5-PE. However, when capturing biohybrid images, the exposure time for each fluorescence was kept at 30 ms.

3.7. GUV imaging and bacteria tracking

Data analysis was performed using the open-source Python software DisGUVery [17], which provides automated detection and quantitative analysis of GUVs in microscopy images. Tracking data were obtained with the TrackMate plugin in ImageJ/Fiji [18]. All graphs and quantitative plots were generated using MATLAB R2022b (MathWorks). The Kruskal-Wallis tests were performed using MATLAB's 'kruskalwallis()' function, with significance determined at a $\alpha = 0.05$.

4. Results and Discussion

4.1. GUV formation

The electroformation of GUVs made of DOPC, cholesterol and DGS-NTA(Ni) lipids were successful. Brightfield and fluorescence microscopy revealed stable, spherical liposomes. As displayed in Figures 2 and 3, the imaging revealed bright, continuous Cy5-PE and GFP fluorescence, indicating even distribution of functional lipids within the DOPC/ cholesterol bilayer.

It was found that the GUVs can survive intact for 2 days at room temperature before bursting. However, despite its stability, several vesicles exhibited multilamellar structures along with some lipid aggregation. The presence of multilamellar GUVs and aggregates may affect overall vesicle population homogeneity and therefore was considered when interpreting size distribution and functional binding analysis.

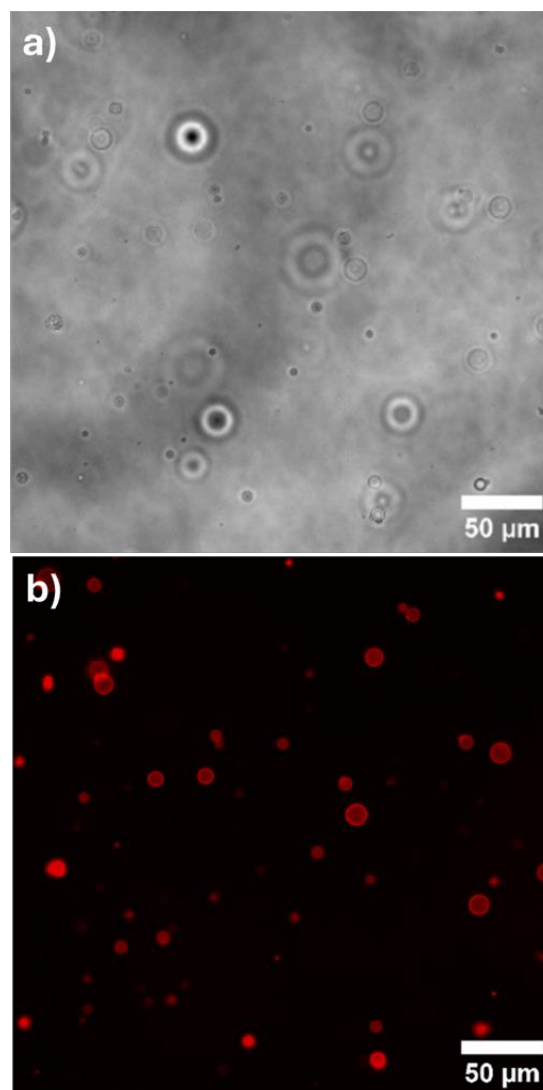


Figure 2. Optical microscope imaging of GUVs at pH 8.0 visualised by a) brightfield and b) Cy5-PE fluorescence

4.2. Effect of 6H-GFP attachment on the GUV membrane due to pH

The electroformed GUVs were incubated in outer solutions with various pH values containing 5 μ M 6H-GFP to optimise the attachment conditions between the GUV membrane containing DGS-NTA(Ni) and the 6-His tags expressed in the GFP.

In Figure 3, it is clear to see that the fluorescence of the liposome membrane is much brighter in pH 8.0 conditions, compared to pH 7.5 and 9.0. However, even in the absence of lipids, the intensity of a fluorescent molecule can vary with pH [10]. Hence, signal-to-background ratios (SBR) were quantified via ImageJ/Fiji [19]: 4.677 ± 1.388 at pH 7.5, 6.682 ± 1.601 at pH 8.0, and 3.197 ± 0.616 at pH 9.0 (Figure 3d). This confirms the peak binding efficiency at pH 8.0. A Kruskal-Wallis test was performed, giving $p = 0.000$, meaning that there

are significant differences between the independent pH groups. Therefore, it is evident that pH does influence the level of GFP attachment.

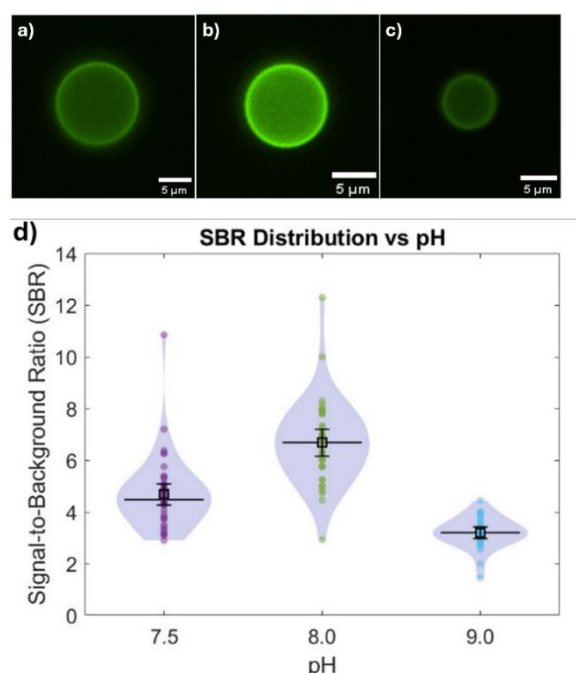


Figure 3. Optical microscope imaging of membrane bound with 6H-GFP at external pH a) 7.5, b) 8.0, and c) 9.0, along with d) the effect of pH on the signal-to-background ratios of GFP-GUV attachment; black square displays mean value with 95% confidence interval, long black line indicates the median values (pH 7.5: 4.48; pH 8.0: 6.69; pH 9.0: 3.20).

The coordinate bonding between the DGS-NTA(Ni) lipid and 6-His tag happens between the Ni^{2+} ion of DGS-NTA(Ni) and the nitrogen of two histidine imidazole rings, where the lone pairs of electron donation comes from the nitrogen [10]. This coordination is pH-dependent because of the pH dependency of the protonation state of His residues [20]. At lower pH, histidine is likely to be protonated, leading to a positive charge [20]. This makes electron donation less probable. Increasing the pH neutralises the histidine, making it a strong metal ion coordinator [20], hence allowing strong binding between its lone electrons and the Ni^{2+} ion of the Ni lipid. However, increasing the pH any higher may lead the histidine to serve as an electron acceptor rather than donor [20]. This suggests why the pH dependence is bell-shaped, peaking at pH 8.0 and reflecting the optimal condition for the His- Ni^{2+} bond.

4.3. GUV size distribution vs pH

Along with analysing the interaction between 6H-GFP and the Ni-lipids embedded in the GUV membrane, the vesicle size distribution for the

varying pH external environment was also quantified (Figure 4) to determine the influence of pH on the GUV morphology.

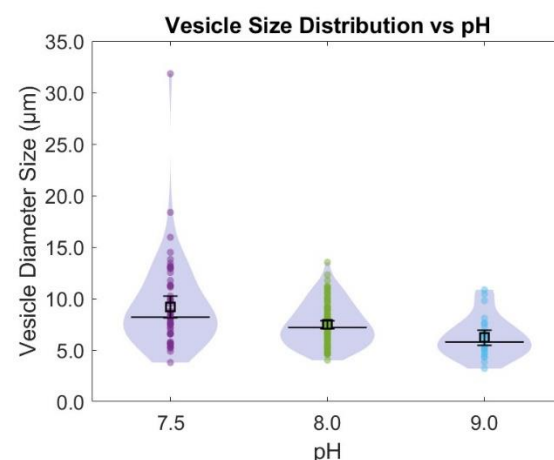


Figure 4. Effect of external pH on GUV diameter distribution. Data points show individual vesicle diameters with black squares representing mean values and 95% confidence intervals; horizontal black lines indicate median values (pH 7.5: 8.21 μm; pH 8.0: 7.19 μm; pH 9.0: 5.78 μm).

The mean vesicle diameter decreased with increasing pH, along with a narrowing size distribution. Incubated at pH 7.5, the mean GUV diameter was found to be 9.22 ± 4.22 μm, with the widest size distribution of 31.83–3.79 μm. At pH 8.0, the mean value decreased to 7.47 ± 2.06 μm with a detected range of 13.52–4.02 μm. The smallest mean value, 6.22 ± 1.95 μm, and the narrowest distribution, 10.86–3.21 μm, was found at pH 9.0. After conducting a Kruskal-Wallis test and Bonferroni-corrected pairwise comparisons, it was revealed that there are great differences between pH 7.5 and 8.0 ($p = 0.0177$), pH 7.5 and 9.0 ($p < 0.0001$), and pH 8.0 and 9.0 ($p = 0.0209$).

With an approximate 33% reduction in mean diameter from pH 7.5 to 9.0, this demonstrates that GUV morphology could be sensitive to the applied external pH conditions. The decrease in range is seen being due to the reduction of the upper bound value. This suggests that the pH-dependent destabilisation primarily affects larger vesicles.

GUV membranes are semipermeable, sharing this characteristic with biological cell membranes [21]. In this study, the pH of the external solution is adjusted with a strong base, NaOH, which dissociates completely in solution. Although the concentration difference of the H^+ ions may not be large enough to cause a significant change to the osmotic pressure, the addition of NaOH when adjusting the pH leads to a more hypertonic outer

solution relative to the internal 0.5 M sucrose solution, hence contributing to osmotic imbalance. This imbalance causes stress on the vesicle membranes, driving H₂O out of the GUVs and reducing internal volume [22]. The increasing external pH leads to larger osmotic stress, leading to vesicle shrinkage and loss of larger structures. As larger vesicles have a bigger surface-to-volume ratio, they are more liable to osmotic stress and hence membrane tension and rupture.

The bursting of liposomes leads to more lipid aggregates in the solution. This could reflect why low GFP intensity on the GUV membrane, Figure 3, is seen for pH 9.0, with 6H-GFP favouring attachment with lipid aggregates instead. Also, smaller GUV size can correlate to there being less surface area for GFP attachment, hence the lower SBR.

4.4. Bacteria motility in different pH buffers

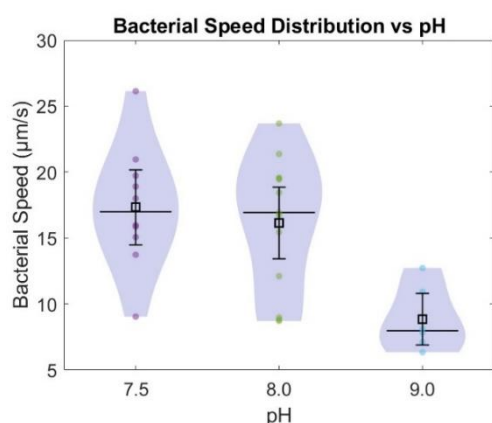


Figure 5. Effect of external pH on the motile *E. coli* speed. Data points show individual vesicle diameters with black squares representing mean values and 95% confidence intervals; horizontal black lines indicate median values (pH 7.5: 16.99 μm/s; pH 8.0: 16.93 μm/s; pH 9.0: 7.98 μm/s).

To find the mean speed of the motile bacteria under different pH environments, the bacteria pellets were resuspended in the different outer solution buffers and observed under an optical microscope in a tunnel slide.

With the mean tracking speeds of the bacteria being 17.341 ± 4.595 μm/s, 16.136 ± 5.016 μm/s and 8.844 ± 2.458 μm/s in ascending order of external pH value (Figure 5), it is indicated that bacteria motility is reduced with increasing pH. Between pH 7.5 and 8.0, the difference is significantly less. Performing the Kruskal-Wallis test, the pairwise comparisons revealed no significant difference between pH 7.5 and 8.0 ($p = 1.0000$), suggesting that mild alkali external environments do not largely

alter motility in this range. However, between pH 7.5 and 8.0, and pH 8.0 and 9.0, significant differences were seen with $p = 0.0069$ and $p = 0.0129$ respectively. This confirms that stronger alkaline conditions reduce swimming speed relative to milder alkaline levels. The non-linear relationship between pH and mean bacteria speed suggests that there could be an upper pH threshold of which the propulsion of motile *E. coli* declines when surpassed.

The motility of bacteria is highly dependent on flagellar rotation, and the flagellar motor is powered by the proton motive force (PMF), which is the flow of protons (H⁺ ions) across the bacterial membrane [23][24]. In previous studies, it is shown that at high pH environments, *E. coli* prioritises its cellular homeostasis above motility energy consumption by increasing proton import mechanisms to maintain its internal cytoplasmic pH of 7.6 [25]. This therefore reduces the energy available for flagellar rotation, decreasing motility. The highest swimming speed in this study at pH 7.5 aligns with the bacteria's native cytoplasmic pH of 7.6, which implies that this is where PMF and motor functions are optimised.

In addition to this, alkaline environments not only influence the flagellar energy expenditure, but it also impacts the flagellar morphology [25][26]. By causing differences in the secondary structure of the polymerised FliC polypeptides that comprised the flagella filaments, high pH environments reduce filament flexibility and impairs coordinated propulsion [26]. Both of these impacts explain the motility threshold effect beyond pH 8.0, impairing effective bacteria movement.

Combining the results found for bacteria speed and GFP-GUV membrane attachment, it was decided not to work any further with pH 7.5 and 9.0 in the next steps of investigating the attachment between the GUVs and motile *E. coli*. With the mean bacteria speed and GFP-GUV attachment being the lowest for pH 9.0, it is shown to be an unideal condition for optimising the transport ability of the motile biohybrids. In the case of eliminating pH 7.5, despite it being the closest to the optimal conditions of *E. coli* motility as discussed above, revisiting Figure 3 displays evidence that GFP-GUV attachment was much more effective in pH 8.0. Also, the difference of the bacterial swim speed between pH 7.5 and 8.0 was found to be insignificant. Hence, these findings suggest that buffering at pH 8.0 maximises both attachment efficiency, vesicle stability and bacteria motility.

4.5. Bacteria-GUV attachment in pH 8.0

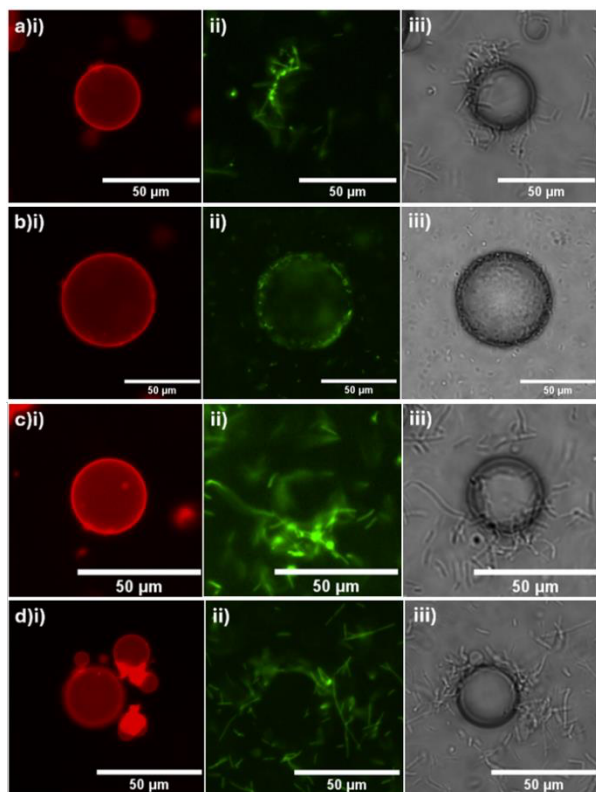


Figure 6. a) *E. coli* expressing Ag-6H, IPTG present; b) different *E. coli* strain expressing Ag-43, but no 6-His tag; c) no IPTG was present during the *E. coli* liquid culture; d) double negative control with no IPTG present in the *E. coli* liquid culture, and no DGS-NTA(Ni) lipids are present in the GUV membrane. Labels i), ii), and iii) represent the Cy5-PE, GFP, and brightfield images in each case respectively. The external solution was 0.21 M glucose in PBS at pH 8.0.

With an external solution of pH 8.0, four circumstances were methodically evaluated to assess the selectivity of biohybrid assembly in regards to Ni^{2+} and His-tag mediated coupling: a) IPTG-induced Ag-6H *E. coli* (6-His surface display, DGS-NTA(Ni) GUVs); b) Ag-43 *E. coli* (no 6-His tag, DGS-NTA(Ni) GUVs); c) no-IPTG Ag-6H *E. coli* (DGS-NTA(Ni) GUVs); and d) double negative control (no IPTG Ag-6H, no DGS-NTA(Ni) lipids), as displayed in Figure 6.

Bacteria pellets were resuspended in 0.21 M glucose in PBS at pH 8.0, which was then used to dilute harvested GUVs (0.5 M sucrose interior) by 100-fold and the mixtures were immediately transferred to microscopy chambers. Attachment and bacteria clustering were tracked over time using an optical fluorescence microscope. The GUV membrane was reported by Cy5-PE, the His-tagged bacterial constructs were reported by GFP, and the bacterial shape and placement in relation to the

vesicle surface were visualised using brightfield imaging.

The images, as shown in Figure 6, displayed bacterial-GUV interactions across all test conditions, which was characterised by bacteria appearing on or near vesicle membranes. In panels a, c and d, the cultured Ag-6H bacteria displayed elongated morphology ($\sim 10 \mu\text{m}$), as well as bacteria clumping on the vesicle surfaces and in solution. Panel b displayed shorter bacteria of a different strain ($\sim 2 \mu\text{m}$) and less pronounced clumping compared to Ag-6H samples. All GUV stock synthesised also had a high presence of lipid aggregates.

In the primary testing (Figure 6a), the GUVs incubated with Ag-6H expressing bacteria exhibited multi-bacterial aggregates rather than uniform single-cell attachment. This could be a contribution of the residual lipid aggregates from GUV formation, providing high density Ni-NTA binding sites that cluster the bacteria. The elongated bacteria, along with the displayed Ag-6H, could also lead to autoaggregation due to increased cell to cell contact [12][27], pointing towards another likely contribution to this clumping.

In contrast to Ag-6H samples, GUVs mixed with untagged motile *E. coli* similarly displayed membrane interactions, despite being with shorter bacteria and less noticeable clumping. This binding implies electrostatic interactions between the positively charged Ni^{2+} ions exposed on DGS-NTA lipids [10] and the negatively charged *E. coli* surface [28].

Figure 6c, displaying no-inducer conditions (lacking 6-His expression), still indicate the presence of bacteria-GUV interactions comparable to induced Ag-6H samples. This suggests that the observed attachment is primarily non-specific, driven by physical trapping in aggregates or electrostatic effects rather than particular Ni^{2+} -His coordination. However, it could be argued that another possibility of this occurrence could be due to the cause of leaky plasmid expressions with the lac repressor not tightly attached, leading to Ag-6H expressions despite the lack of inducer. [29][30].

Figure 6d, also non-induced but this time without the presence of Ni-lipids, displayed similar levels of bacterial contact after 30 min. However, this association was shown substantially faster with Ni-NTA present. Bacteria were visible on membranes in around 5 mins opposed to 30 mins for no-Ni controls. Since there are no particular binding

sites for directed attachment, the lack of functional lipids indicates that lipid aggregation from the electroformation process, rather than Ni-NTA incorporation, is a significant cause of bacterial entrapment.

These controls reveal that current GUV preparations are negatively hindered by significant lipid aggregation, trapping bacteria non-specifically and obscuring genuine Ni^{2+} -His attachments. Although still complicated by aggregates, the accelerated attraction kinetics with Ni-NTA GUVs (~5 mins vs 30 mins) offer the strongest evidence for an electrostatic contribution from Ni^{2+} ions interacting with the negatively charged bacterial surfaces, assuming the plasmid expressions were not leaky.

5. Conclusions and Outlooks

This study establishes an optimal pH influencing motile biohybrid assembly between 6-His tagged *E. coli* and DGS-NTA(Ni) integrated, DOPC based GUVs. It was revealed that an external solution of pH 8.0 represents the optimal condition, balancing average *E. coli* motility speed (16.136 $\mu\text{m/s}$), Ni^{2+} -His attachment efficiency (SBR 6.682) and the structural integrity of the electroformed GUVs. However, widespread lipid aggregates and nonspecific electrostatic interactions between the negatively charged bacteria membrane and Ni^{2+} ions present in the DGS-NTA(Ni) lipids, evident across all tested controls, show that the present lipid mixture composition and electroformation technique used conflict precise Ni^{2+} -His coupling with physical entrapment, calling for necessary modifications in vesicle creation.

For future studies, it is important to eliminate the presence of lipid aggregates. This could be done by purifying the GUVs or getting rid of lipid aggregates by using microfluidic devices [31] as well as optimising the total lipid concentration during GUV preparation. In addition, evaluating different GUV synthesis methods should also be further explored, as previous reports have indicated that electroformation on platinum wire or PVA hydrogel swelling lead to improved His-tagged GFP binding compared to electroformation via ITO slides [10]. Although the microscopic imaging suggests potential bacteria-GUV interactions due to evident close proximity, dedicated examinations are needed to unambiguously confirm bacterial attachment to the GUV surface, such as 3D confocal colocalization [32].

Ultimately, by resolving aggregation artifacts, this motile biohybrid assembly method could enable great advancement to targeted, non-invasive drug delivery. The strong yet reversible Ni^{2+} -His bond allows the bacteria to transport drug-loaded liposomes and release them at target sites in response to various stimuli, such as changes in pH. It also further supports studies working with biohybrids as it provides a simplified one-step assembly platform, eliminating the need of additional binding lipids.

Acknowledgements

I would like to extend my gratitude to Dr. Jyoti P. Gurung for his invaluable help throughout the project, including performing the bacteria transformation, PhD student Joshua Christopher for preparing and providing the motile *E. coli* bacteria strain used, and Dr. Marcus Fletcher for providing the 6H-GFP. I am also very grateful to all members of the Elani Group for their continuous support.

References

- [1] Zhang, Y., Obuchi, H. and Toyota, T. (2023) 'A practical guide to preparation and applications of giant unilamellar vesicles formed via centrifugation of Water-in-Oil emulsion droplets,' *Membranes*, 13(4), p. 440. <https://doi.org/10.3390/membranes13040440>.
- [2] Parigoris, E. *et al.* (2020) 'Facile generation of giant unilamellar vesicles using polyacrylamide gels,' *Scientific Reports*, 10(1), p. 4824. <https://doi.org/10.1038/s41598-020-61655-2>.
- [3] Staufer, O. *et al.* (2020). 'Microfluidic production and characterization of biofunctionalized giant unilamellar vesicles for targeted intracellular cargo delivery,' *Biomaterials*, 264(27), p. 120203. <https://doi.org/10.1016/j.biomaterials.2020.120203>.
- [4] Crosio, M. A. *et al.* (2024). 'Giant unilamellar vesicles (GUVs): A key tool for the study of biophysical properties of membranes,' *Advances in Biomembranes and Lipid Self-Assembly*, 40, p. 55-65. <https://doi.org/10.1016/bs.abl.2024.10.005>.
- [5] Bally, M. *et al.* (2010) 'Liposome and lipid bilayer arrays towards biosensing applications,' *Small*, 6(22), pp. 2481–2497. <https://doi.org/10.1002/sml.201000644>.
- [6] Nair, K. S. and Bajaj, H. (2023) 'Advances in giant unilamellar vesicle preparation techniques and

applications,' *Advances in Colloid and Interface Science*, 318, p. 102935. <https://doi.org/10.1016/j.cis.2023.102935>.

[7] Daraee, H. et al. (2014) 'Application of liposomes in medicine and drug delivery,' *Artificial Cells Nanomedicine and Biotechnology*, 44(1), pp. 381–391. <https://doi.org/10.3109/21691401.2014.953633>.

[8] Nagard, L.L. et al. (2022) 'Encapsulated bacteria deform lipid vesicles into flagellated swimmers,' *Proceedings of the National Academy of Sciences*, 119(34), p. e2206096119. <https://doi.org/10.1073/pnas.2206096119>.

[9] Dogra, N., Izadi, H. and Vanderlick, T.K. (2016) 'Micro-motors: A motile bacteria based system for liposome cargo transport,' *Scientific Reports*, 6(1), p. 29369. <https://doi.org/10.1038/srep29369>.

[10] Pramanik, S. et al. (2022) 'Binding of His-tagged fluorophores to lipid bilayers of giant vesicles,' *Soft Matter*, 18(34), pp. 6372–6383. <https://doi.org/10.1039/d2sm00915c>.

[11] Bastos-Arrieta, J. et al. (2018) 'Bacterial biohybrid microswimmers,' *Frontiers in Robotics and AI*, 5, p. 97. <https://doi.org/10.3389/frobt.2018.00097>.

[12] Schauer, O. et al. (2018) 'Motility and chemotaxis of bacteria-driven microswimmers fabricated using antigen 43-mediated biotin display,' *Scientific Reports*, 8(1), p. 9801. <https://doi.org/10.1038/s41598-018-28102-9>.

[13] Xiao, Q. et al. (2022) 'Co-assembly of liposomes, Dendrimersomes, and Polymersomes with amphiphilic Janus dendrimers conjugated to Mono- and Tris-Nitrilotriacetic Acid (NTA, TrisNTA) enhances protein recruitment,' *Giant*, 9, p. 100089. <https://doi.org/10.1016/J.GIANT.2021.100089>

[14] Currie, S. (2025) *What's the difference between nickel NTA and nickel IDA agarose beads?* <https://www.goldbio.com/blogs/articles/whats-the-difference-between-nickel-nta-and-ida> (Accessed: November 30, 2025).

[15] *FluoroFinder* (no date). <https://app.fluorofinder.com/dyes/82-gfp-ex-max-488-nm-em-max-510-nm> (Accessed: November 25, 2025).

[16] *FluoroFinder* (no date b). <https://app.fluorofinder.com/dyes/22-cy5-ex-max-649-nm-em-max-667-nm> (Accessed: November 25, 2025).

[17] Van Buren, L., Koenderink, G.H. and Martinez-Torres, C. (2022) 'DisGUVery: a versatile Open-Source software for High-Throughput image analysis of giant unilamellar vesicles,' *ACS Synthetic Biology*, 12(1), pp. 120–135. <https://doi.org/10.1021/acssynbio.2c00407>.

[18] Ershov, D. et al. (2022) 'TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines,' *Nature Methods*, 19(7), pp. 829–832. <https://doi.org/10.1038/s41592-022-01507-1>.

[19] Schindelin, J. et al. (2012) 'Fiji: an open-source platform for biological-image analysis,' *Nature Methods*, 9(7), pp. 676–682. <https://doi.org/10.1038/nmeth.2019>.

[20] Calinsky, R. and Levy, Y. (2024) 'Histidine in Proteins: pH-Dependent Interplay between π - π , Cation- π , and CH- π Interactions,' *Journal of Chemical Theory and Computation*, 20(15), pp. 6930–6945. <https://doi.org/10.1021/acs.jctc.4c00606>.

[21] Toyota, T. and Zhang, Y. (2022) 'Identifying and Manipulating Giant Vesicles: Review of Recent Approaches,' *Micromachines*, 13(5), p. 644. <https://doi.org/10.3390/mi13050644>.

[22] Liu, X. et al. (2022) 'Vesicles Balance Osmotic Stress with Bending Energy That Can Be Released to Form Daughter Vesicles,' *The Journal of Physical Chemistry Letters*, 13(2), pp. 498–507. <https://doi.org/10.1021/acs.jpclett.1c03369>.

[23] Morimoto, Y. and Minamino, T. (2014) 'Structure and function of the Bi-Directional Bacterial Flagellar Motor,' *Biomolecules*, 4(1), pp. 217–234. <https://doi.org/10.3390/biom4010217>.

[24] Nakamura, S. and Minamino, T. (2024) 'Structure and dynamics of the Bacterial Flagellar Motor Complex,' *Biomolecules*, 14(12), p. 1488. <https://doi.org/10.3390/biom14121488>.

[25] Maurer, L.M. et al. (2004) 'pH Regulates Genes for Flagellar Motility, Catabolism, and Oxidative Stress in Escherichia coli K-12,' *Journal of Bacteriology*, 187(1), pp. 304–319. <https://doi.org/10.1128/jb.187.1.304-319.2005>.

- [26] Cheng, S. J. *et al.* (2003) 'Nanoscope analysis on pH induced morphological changes of flagella in *Escherichia coli*,' *Journal of Microbiology, Immunology and Infection*, 46(6), pp. 405–412. <https://doi.org/10.1016/J.JMII.2012.08.004>.
- [27] Trunk, T., Khalil, H.S. and Leo, J.C. (2018) 'Bacterial autoaggregation,' *AIMS Microbiology*, 4(1), pp. 140–164. <https://doi.org/10.3934/microbiol.2018.1.140>.
- [28] Liang, X. *et al.* (2016) 'E. coli Surface Properties Differ between Stream Water and Sediment Environments,' *Frontiers in Microbiology*, 7, p. 1732. <https://doi.org/10.3389/fmicb.2016.01732>.
- [29] Kroemer, T. (2020) *How does IPTG induction work?* <https://www.goldbio.com/blogs/articles/how-does-iptg-induction-work?srsltid=AfmBOopfULJ40V5DXzfwIVAfzPqEZMOPjCmAfcjSsHgb7ogVT5WQ2IU0> (Accessed: December 3, 2025).
- [30] Singh, M.I. and Jain, V. (2013) 'Tagging the Expressed Protein with 6 Histidines: Rapid Cloning of an Amplicon with Three Options,' *PLoS ONE*, 8(5), p. e63922. <https://doi.org/10.1371/journal.pone.0063922>.
- [31] Tivony, R. *et al.* (2021) 'A microfluidic platform for sequential assembly and separation of synthetic cell models,' *ACS Synthetic Biology*, 10(11), pp. 3105–3116. <https://doi.org/10.1021/acssynbio.1c00371>.
- [32] Bakar, Z.H.A. *et al.* (2023) 'Optimization of 3D immunofluorescence analysis and visualization using IMARIS and MeshLab,' *Cells*, 12(2), p. 218. <https://doi.org/10.3390/cells12020218>.

Fabrication of Composite SU-102@Anodic Aluminium Oxide Membranes with Characterisation and Anionic Dye Rejection Properties

Xinyue Xu and Yu Qian

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Two main targets for membrane nanofiltration are achieving high permeance and rejection, but there is a trade-off between them. To maintain the high potential for water treatment, this study analyzes two performance metrics for the SU-102 metal-organic framework. Zirconium-ellagate MOF (SU-102) membranes were synthesized by in situ growth on anodic aluminium oxide (AAO) substrates with 20 nm and 200 nm pore sizes to tune transport pathways. Pure-water permeance stabilised at 410 and 780 LMH bar⁻¹ for the 20 nm and 200 nm substrates, respectively. This is consistent with reduced hydraulic resistance on the larger-pore substrate. Rejection of anionic dyes with increasing size showed steric sieving, which was further enhanced by electrostatic exclusion. At a similar molecular weight, Allura Red, with a formal charge of -2, exhibited higher rejection than Orange G with a formal charge of -1. This study indicates that SU-102 membranes exhibit excellent performance due to their high hydrophilicity, negatively charged surface, and tunable pores, depending on the substrate pore size adopted. Each contributes to good water transport, size-sieving, and electrostatic exclusion selectivity.

Keywords: Metal-Organic Framework, SU-102, Membrane, Substrate, Nanofiltration, Dye Removal, Anodic Aluminium Oxide Substrate, Permeation, Rejection

1. Introduction

Nowadays, due to severe water pollution and a rapidly growing population, water treatment is becoming increasingly important. Many contaminants, such as pharmaceutical compounds, dyes, and pesticides, are discharged directly into the water cycle rather than being adequately treated. This inadequate treatment severely damages ecosystems and human health [1]. Among all these contaminants, synthetic dyes stand out for their aromatic structures, strong colour intensity, and chemical stability. These characteristics make them very difficult to decompose naturally, so they remain quite harmful even at low concentrations [2, 3].

Separation by membrane offers a potential strategy because it can be easily scaled from lab-scale to industrial-scale and provides benefits in energy efficiency, low cost, and long-term stability. It can also provide controllable selectivity by tuning its structures and will not form secondary contaminants afterwards [4]. For separating low-molecular-weight dyes, nanofiltration is preferable to reverse osmosis because it requires much lower operating pressure [5].

These factors emphasize the requirement of a membrane for high stability and selective separation. Among different materials, the zirconium-based metal-organic frameworks show a high potential. From the structure, the strong zirconium-oxygen bonding can withstand severe chemical and thermal conditions. Also, for the coordination part, highly connected zirconium metal ions with carboxylate groups can provide stable water permeability, resist fouling, chemical erosion, and structural decomposition over a wide pH range.

The clear, tunable pore structure enables selective charge and size separation during the separation process. The high surface area of the porous structure also provides many coordination points for interaction with the solute. Overall, these features make zirconium MOF attractive for future studies [6]. From past research, zirconium-based MOF has already shown a great adsorption ability to various pollutants, such as pharmaceuticals and personal care products (PPCPs) [7], polyfluoroalkyl substances (PFAS) [8], organic dyes [9, 10, 11], heavy-metal ions [12, 13], and radionuclides [14, 15].

Zr-MOFs also exhibit good compatibility, enabling uniform, well-distributed growth on inorganic and polymer substrates [16]. As a result, many scientists have tried to use this nanostructure to separate dyes. For example, MOF-808 can reject both cationic and anionic dye molecules by about 99%. It can maintain this high rejection rate for operating over 96 hours [17]. Similarly, UiO-66/PGP composite membranes can achieve rejection rates over 94% for charged dye molecules and antibiotics. It also shows good fouling resistance and long-term stability [18]. Therefore, all these previous studies show that Zr-MOFs have great potential for future investigations on separating pollutants and dyes.

Although there is extensive prior research on Zr-MOFs, studies on the SU-102 framework remain limited, which has unique advantages. However, SU-102 was first reported by the Stockholm University team, based on studies by Grape et al.'s group [19]. It combines ellagic acid as the organic ligand part and Zr(IV) metal ions. Ellagic acid is selected because it is a sustainable, renewable linker that can

be naturally sourced from tree bark and fruit peels. As a result, it also benefits from a cost perspective. This non-toxic, green material contains many hydroxyl and ester groups. These coordination points can bond strongly with the metal part to form a stable framework. Compared to UiO-66, SU-102 offers a more sustainable approach, as UiO-66 relies mainly on the petrochemical, non-renewable BDC linker [20]. The SU-102 combination provides a naturally negatively charged framework, thereby exhibiting strong electrostatic attraction to positively charged organic molecules. Grape et al.'s group has already shown that they can effectively remove a wide range of positively charged emerging organic contaminants (EOCs) from real-world industrial wastewater effluent. Their process can achieve a rejection rate of 79-99.6%. This study shows that electrostatic interactions can play an important role. But UiO-66 and MOF-88 are neutral, so they cannot separate contaminants based on charge without introducing modifications to their original, naturally grown structures [21].

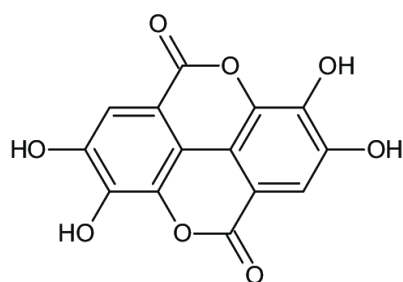


Figure 1.1 Ellagic Acid Molecule Skeletal Formula

Given the original structural stability of Zr-MOFs, using ellagic acid as the ligand can further improve stability, as it has multidentate coordination. Therefore, it can retain the good crystallinity in the real wastewater effluent. This robustness is a significant advantage over MIL-100 and other related carboxylate MOFs, which usually collapse and decompose during hydrolysis under aqueous conditions [22]. Moreover, SU-102 contains several hydroxyl and ester groups, making it naturally hydrophilic and promoting water transport. However, UiO-66 must be functionalized with hydrophilic groups, such as amines or hydroxyls, to increase its permeance [23]. From a structure perspective, SU-102 has well-defined one-dimensional honeycomb channels with a pore size of 1.2nm. This intrinsically porous formwork further realizes the effective size sieving and fast mass transport.

In summary, SU-102 has several mechanisms that make separation efficient: its intrinsic negative surface charge, its highly hydrophilic nature, and tunable pathway sizes. Therefore, when rejecting specific contaminants, the Donnan Exclusion by charges, molecular size-sieving, steric hindrance,

and adsorption can all contribute to a high selectivity. To further assess its charge-selectivity and steric-hindrance mechanism, SU-102@AAO membranes were evaluated with negatively charged dyes, where electrostatic repulsion reduces adsorption, allowing the observed rejection to reflect primarily molecular-size exclusion. AAO was selected as the substrate because its uniform, hydroxyl-rich surface promotes controlled MOF nucleation [24], and its tunable pore structure allows examination of substrate effects. Specifically, AAO substrates with 20 nm and 200 nm pores provide distinct environments that influence precursor infiltration, MOF growth, and membrane structure and performance.

3. Methods

3.1 Materials

Zirconyl chloride octahydrate (white crystalline solid, 98%, metal basis) was purchased from Alfa Aesar, and ellagic acid (pale yellow powder) was obtained from Thermo Fisher Scientific. Methanol (100%), glacial acetic acid (97%), and N, N-dimethylformamide (DMF) used as solvents were sourced from VWR. Anodic aluminium oxide (AAO) substrates with a diameter of 25 mm were supplied by Cytiva, and Teflon-lined stainless-steel autoclaves were acquired from Parr Instrument Company. For membrane performance evaluation, two dead-end filtration cells were employed: a commercial Sterlitech HP4750 unit and a custom-built filtration apparatus fabricated in the Mechanical Engineering workshop. The dyes Orange G and Rose Bengal were purchased from Thermo Fisher Scientific, whereas Allura Red, Congo Red, and Methyl Blue were obtained from Sigma-Aldrich. The details of these dyes are attached in Section 9: Supplementary Information. Centrifuge tubes and microcentrifuge tubes are used to collect permeate and retentate were sourced from VWR.

3.2 Fabrication of SU-102@AAO Membranes

SU-102@AAO membranes were synthesized on anodic aluminium oxide (AAO) substrates using a capillary-assisted in situ growth method. A precursor solution was prepared by dissolving ellagic acid (340 mg) and zirconyl chloride octahydrate ($\text{ZrOCl}_2 \cdot 8\text{H}_2\text{O}$, 182 mg) in 45 mL of a mixed solvent of N, N-dimethylformamide (DMF), acetic acid, and deionized water (2:1.5:1 v/v). The AAO substrates were pre-immersed in the precursor solution at room temperature to promote capillary infiltration, then transferred with the solution into a Teflon-lined stainless-steel autoclave and placed in a preheated oven at 100°C for 24h. The resulting SU-102@AAO composite membranes were then thoroughly rinsed with methanol to eliminate residual reagents and solvents and dried at 80 °C overnight.

3.3 Methodology for Water Permeation

This study investigates the influence of substrate pore sizes (20 nm & 200 nm) on the performance of the SU-102@AAO membrane, with particular emphasis on water permeance and dye rejection. B Ultrasonic treatment is taken before each experiment to clean the components and remove residual contaminants. Under constant transmembrane pressure, pour approximately 150 mL of DI water into the cell, and then perform the dead-end separation experiment. Filtration was considered to begin upon the appearance of the first permeate droplet. Permeate was collected in centrifuge tubes, and cumulative permeate mass was recorded over time to calculate the water permeance. Measurements were taken at 5-minute intervals to reduce short-term fluctuations in the permeance data.

The membrane permeance (A) was determined from the permeate flux (J) using the following relationships. The permeate flux was calculated as:

$$J = \frac{\Delta V}{A \cdot \Delta t} \quad (1)$$

Where J is the permeate flux in LMH , ΔV is the change in permeate volume in L , A is the effective membrane area in m^2 , and Δt is the sampling time interval in h .

The membrane permeance was subsequently obtained from:

$$A = \frac{J}{\Delta P} \quad (2)$$

Where A is the permeance in $LMH \text{ bar}^{-1}$, ΔP is the applied transmembrane pressure in bar .

3.4 Methodology for Dye Rejection Performance

For each dye filtration test, the procedure described in Section 3.3 was followed, except that DI water was replaced with the corresponding dye solutions. Also, a magnetic rotation of 300 rpm was set to avoid concentration polarization. To assess the effect of molecular size on separation performance, a series of dyes with increasing molecular weight were filtered in sequence from smallest to largest. Before introducing each new dye solution, the membrane module and the cell were thoroughly flushed with deionised water to remove any residual dye or contaminants. Flushing continued until the permeate, analysed using a NanoDrop UV-Vis spectrophotometer, produced an absorbance spectrum, which was compared with DI water as the blank to confirm system baseline stability.

A feed sample was collected before filtration to determine the initial dye concentration. Residual flushing water in the tube was discarded. The dead-end cell was alternately pressurised and depressurised at 30-minute intervals for the first four hours to enable periodic sampling of both permeate and retentate. During the next six hours, the

pressurisation cycle was reduced to once per hour. This higher sampling frequency early on provided better temporal resolution. It captured both initial adsorption and later membrane separation behaviour.

Before each depressurisation step, permeate was collected into microcentrifuge tubes for approximately 10 minutes to ensure a sufficient sample volume, prevent concentration gradients, and improve accuracy. The gas valve was then closed to release the internal pressure. The cell was opened, and retentate was collected from the top, middle, and bottom of the cell to obtain a representative composite sample of approximately 1 mL. After 10 hours of operation, the system was left undisturbed until the 24-hour mark, after which samples were collected for long-term performance evaluation. Continuous permeate was collected in centrifuge tubes for volume measurement, and permeance was calculated following the method described in Section 3.3. Since the dye solutions used in this study are 20 ppm, which is sufficiently dilute that their density was effectively indistinguishable from that of pure water, the density of deionised water was adopted for all dye permeance calculations.

3.5 Methodology for UV-Vis Analysis

Feed, permeate, and retentate samples were analysed using a NanoDrop UV-Vis spectrophotometer to quantify dye rejection. Before every analysis, the detector surface was cleaned sequentially with a wet and a dry laboratory wipe, and fresh deionised (DI) water was used for both the blank and the rinsing steps to avoid cross-contamination. Each sample was measured in triplicate, and the average absorbance value was used to minimise instrumental variability. Samples that could not be analysed on the same day were sealed with parafilm and stored in a low temperature environment to minimise degradation.

Rejections relative to the feed and retentate were calculated using the equations provided below. In cases where filtration was paused and continued the following day, a fresh feed sample was collected before resuming the experiment to ensure accurate rejection determination.

$$R_f = 1 - \frac{C_p}{C_f} \quad (3)$$

$$R_r = 1 - \frac{C_p}{C_r} \quad (4)$$

Where R_f , R_r are the rejections calculated with respect to the feed and retentate, C_p , C_f , C_r are the concentrations of the permeate, feed, and retentate solutions. According to the fundamental principle in analytical chemistry, also known as Beer's Law [25], the absorbance of a solution is directly proportional to the concentration of the solute. Therefore, UV-

Vis absorbance is used to measure rejection, since the concentration ratio equals the absorbance ratio.

3.6 Characterisation of SU-102@AAO Membrane

The fabricated SU-102@AAO membranes were then systematically characterised to assess their morphology, crystallinity, hydrophilicity, framework formation, and thermal stability. The corresponding techniques are Scanning Electron Microscopy (SEM), Contact Angle, Fourier-Transform Infrared Spectroscopy (FTIR), X-Ray Diffraction (XRD), and Thermogravimetric Analysis (TGA). Detailed descriptions and results for each technique will be presented below.

4. Results

4.1 Scanning Electron Microscopy (SEM)

Scanning electron microscopy (SEM) was employed to quantify surface morphology, grain and defect density, film continuity, and cross-sectional thickness of the SU-102@AAO layer. A continuous, defect-free, uniform-thickness film minimizes bypass and pinhole leakage. This improves solute rejection. A controlled microstructure and low roughness can reduce foulant deposition. Thickness and morphology also dictate hydraulic resistance and thus water permeance. Consequently, smoother, more continuous SU-102@AAO membranes deliver higher flux at a fixed pressure while maintaining size-sieving selectivity.

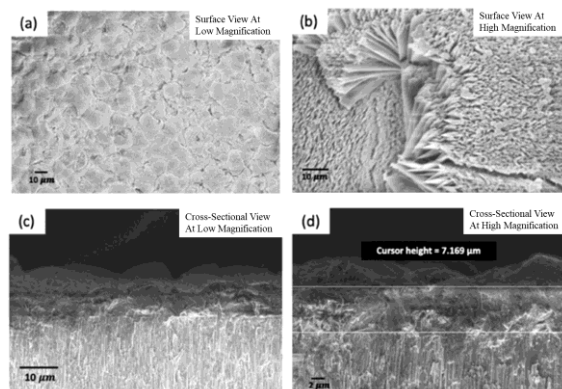


Figure 4.1 SEM Images of SU-102@AAO Membrane: (a,b) Surface Views and (c, d) Cross-Sectional View at Low and High Magnification.

Figure 3.2.1 demonstrates features that support successful infiltration and surface nucleation during the capillary-assisted in situ process on the AAO substrate. The picture (a) shows uniform coverage with a rough, granular texture indicative of densely packed microcrystallites, while the higher magnification surface view (b) reveals micron-scale spherulitic domains that nucleated at discrete sites and grew radially. The pathways grow vertically aligned with the substrate, providing potential for water permeance. The cross-sections (c) display a continuous SU-102@AAO layer conformally attached to the AAO substrate. The picture (d)

shows a compact top layer and intergrown crystallites bridging the pore mouths, yielding a film approximately 7.17 μm thick. This morphology is consistent with pre-immersion-driven capillary infiltration of the precursor, crystallization at 100 $^{\circ}\text{C}$ for 24 hours, and post-synthesis methanol rinsing and drying at 80 $^{\circ}\text{C}$, which together produce a continuous, defect-minimized membrane.

4.2 Contact Angle Test

Hydrophilicity was quantified by measuring the equilibrium water contact angle (θ). When $\theta < 90^{\circ}$, the membrane is hydrophilic, and when $\theta > 90^{\circ}$, it is hydrophobic. A low θ indicates substantial wetting and high interfacial affinity. This will suppress the adsorption of hydrophobic foulants and accelerate water uptake. At the same time, greater wettability also stabilises solute rejection by minimizing dye adsorption and increasing water transport.

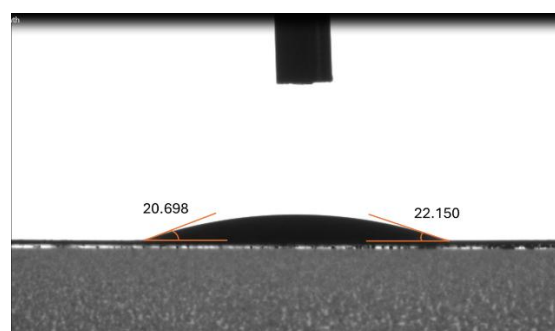


Figure 4.3 Water Contact Angle on Membrane Surface

From Figure 3.2.3, the left and right baseline angles are $\theta_L = 20.70^{\circ}$ and $\theta_R = 22.15^{\circ}$, yielding an average contact angle of $\theta = 21.43^{\circ}$. The low angle indicates a highly hydrophilic surface. The droplet spreads rapidly and maintains a low profile, consistent with high surface affinity for water. Slight left-correct asymmetry arises from local porosity and substrate heterogeneity.

4.3 Fourier-Transform Infrared Spectroscopy (FTIR)

FTIR spectroscopy was used to identify the chemical bonding and functional groups in the SU-102@AAO membrane. For this membrane, FTIR is used to focus on aromatic C=C stretches, carboxylate COO⁻ vibrations in both symmetric (ν_s) and asymmetric (ν_{as}) modes, and metal-ligand modes such as Zr-O-C, Zr-O, and Zr-OH. These features are very critical because they confirm that the ellagic-acid linker sites bond correctly to Zr to form the desired metal-organic framework. Observation of distinct ν_s and ν_{as} carboxylate bands with an intermediate separation $\Delta\nu = \nu_{as} - \nu_s$ confirms that ellagic-acid linkers coordinate to zirconium nodes in a bridging bidentate mode. This observation is essential because it verifies the connectivity of SU-102@AAO's framework and its

net negative surface charge. The intensity of the Zr-OH and O-H envelope reflects the abundance of hydrophilic sites, indicating enhanced water uptake and transport. Moreover, the presence of continuous Zr-O-C networks and carboxylates supports steric -sieving and a negative surface charge. As a result, testing with anionic dyes reveals pore-size-based rejection, and hydrophilic groups further support high water permeability and stable separation performance.

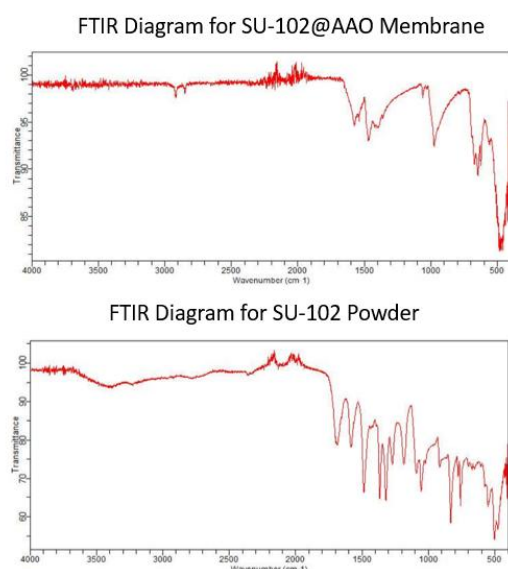


Figure 4.4 FTIR Diagram for the SU-102@AAO Membrane and SU-102 Powder

Figure 4.4 provides diagrams of the SU-102@AAO membrane and SU-102 powder under identical instrument settings. Both exhibit characteristic vibrational modes confirming the successful coordination of the ellagate linker to the zirconium. However, compared to the powder spectrum, the membrane spectrum shows lower-signal-to-noise bands. Since the SU-102 layer on the AAO substrate is very thin and brittle, it will limit the optical path length and result in imperfect contact with the transmission interface of the FTIR machine detector. Also, the AAO substrate will dilute the SU-102 mass within the detected area. In contrast, the centrifuged powder provides higher mass per unit area and intimate contact. This can provide sharper, more intense peaks. For powder preparation, after in situ growth, the AAO substrate was immersed in a precursor solution and reacted at 100 °C for 24 hours; the resulting suspension was then centrifuged at 6000 rpm. The solids collected from the centrifuge were dried at 80 °C overnight, which is the same as the membrane post-treatment. Although the membrane spectrum shows reduced intensity, it still displays diagnostic bands, confirming framework formation on the AAO substrate.

From the diagram of powder, there are several important peaks that confirm the successful formation of the framework. First, at approximately 1600 cm⁻¹ and 1500 cm⁻¹, the aromatic C=C stretching occurs. At 1590-1600 cm⁻¹ and 1400 cm⁻¹, carboxylate asymmetric [$\nu_{as}(\text{COO}^-)$] and symmetric [$\nu_s(\text{COO}^-)$] stretches happen. The two peaks confirm a bidentate coordination to the metal ion, which further enhances the stability of the framework. To further assess how the organic ligand part is bonded with the metal, Zr-O-C peaks should be analysed. From the diagram, C-O stretching vibrations are observed at 1280 cm⁻¹, 1210 cm⁻¹, and in the 1110-1050 cm⁻¹ region. Zirconium with oxo cluster happens at low-frequency range 670 and 520 cm⁻¹. In order to support the ellagate backbone structure, the aromatic out-of-plane deformation shows a peak in the 872-744 cm⁻¹ range. Moreover, in the 3600-3000 cm⁻¹ region. The broad spanning peak is the O-H stretching vibrations. This shows the Zr-OH groups and hydrogen-bonded water molecules. This supports the hydrophilic character of SU-102.

4.4 X-Ray Diffraction (XRD)

The X-ray Diffraction (XRD) technique was applied to compare the simulated SU-102 crystal and the experimental one, focusing on the crystallinity, phase purity, and certain patterns. These features are used to confirm that the SU-102 can reach the expected pore size for the structure and size-sieving ability. As a result, if the phase is pure, highly crystallized, and shows the correct orientation, it can provide preferred porosity. Therefore, it can provide steric hindrance and still maintain a stable and acceptable water transport.

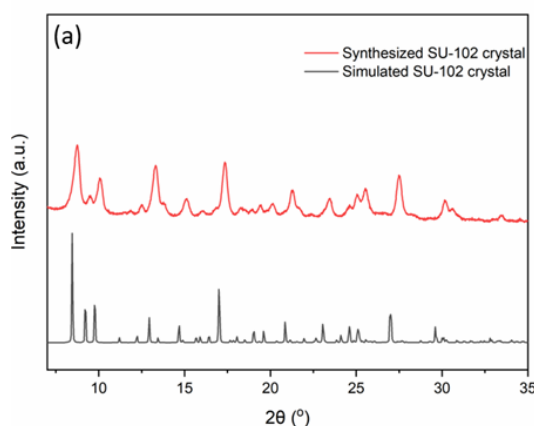


Figure 4.2 XRD Image of SU-102 Powder: Experimental Pattern and Simulated Reference

From Figure 4.2, in the range of $2\theta = 5^\circ - 35^\circ$, both lines show a high similarity, which means the synthesized crystal closely matches closely to the expected pattern. This high similarity is evidence for the successful formation of the intended framework and good crystallinity. The experimental peaks are wider due to the limited coherent domain size. The

differences in relative intensities are due to the preferred crystallite orientation. Also, there are no additional reflections in the detection limit, so the phase is pure.

4.5 Thermogravimetric Analysis (TGA)

Thermogravimetric analysis (TGA) was used to quantify how the membrane mass changes during the pyrolytic decomposition. This technique was adopted because it can critically show the thermal stability of the framework, especially under high-temperature operations. For Stage 1, volatiles like physically adsorbed water molecules, acetic acid solvent, and methanol used to rinse the membrane as the post-treatment will be removed at this low temperature regime. Hence, the first stage defines an activation startup that clears the pores without damaging the framework. This can maximise the pore accessibility and increase the water flux. Stage 2 falls within an intermediate-temperature region. This stage reflects the dehydroxylation of zirconium and loss of more strongly bound guests. This regime measures the start of structural weakening. At Stage 3, in the higher-temperature region, the organic ellagate linkers decompose. The decomposition leaves a stable inorganic residue and indicates the proportion of inorganic fraction of the material that can withstand high temperatures. Combining all three stages, the TGA diagram can quantify the thermal robustness of membranes that tolerate higher temperatures without significant mass loss. These features help preserve membrane structure and rejection performance during long-term operation at high temperatures.

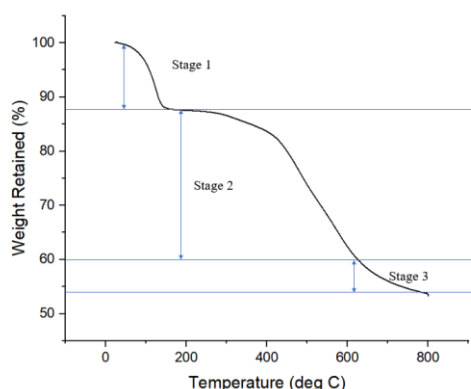


Figure 4.4 TGA Graph of SU-102@AAO Membrane in N₂ with 40mLmin⁻¹ and Heat at 10°C min⁻¹

Figure 4.4 shows the thermal decomposition mechanism of the SU-102@AAO membrane with N₂ injected at 40 mL per minute, and a heat rate of 10 °C min⁻¹. The figure shows three distinct stages. For the first stage, where the temperature is from 30 to 200 °C, the change in mass is approximately 12%. This mass change is due to the desorption of weakly bound guest molecules, such as physisorbed H₂O, residual synthesis solvents, and methanol used to

wash the membrane. For the second stage, the temperature is 200-650 °C. During this stage, the major mass loss occurs, which is approximately 28%. This high mass loss is due to pyrolytic decomposition of the SU-102 metal-organic framework, including the linker scission and node dehydroxylation. Since nitrogen gas was used, the oxidation reaction did not affect the result. In the third stage, where the temperature is 640-800 °C, the change in mass is only 6%. In this stage, a stable residue is formed through gradual carbonization and the expulsion of remaining volatiles.

Also, due to the injection of N₂ rather than air, the process prevents complete organic combustion, yielding a final residue with a mass percentage of 54%, including the AAO substrate and zirconium-based inorganic species. The cumulative mass loss across the entire temperature regime is about 46%.

4.6 Water Permeation Performance

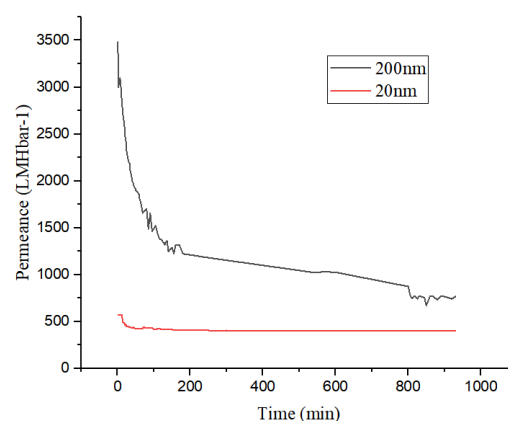


Figure 4.6 Water Permeation Test for SU-102@AAO Membrane

From Figure 4.6, the SU-102@AAO membrane grown on the 200 nm substrate exhibits a much larger water permeance than the SU-102@AAO membrane grown on the 20 nm substrate. This is due to a higher flux through more open hydraulic pathways enabled by the substrate's larger pore size. Initially, the 200 nm sample shows a high permeance at 3000 LMH bar⁻¹, which decays to approximately 900 LMH bar⁻¹ by about 200 min and then stabilizes near 780 LMH bar⁻¹ by about 1000 min. This decrease is consistent with the increasing wetting of the framework and mechanical compaction. Once the experiment starts and a constant pressure is applied, the initially opened intercrystalline gas will gradually close, and the bypass flow will also be eliminated. In contrast, the 20 nm substrate yields a lower but immediately stable permeance of 410 LMH bar⁻¹. This observation indicates that creating more effective, narrower, double-layer overlapped passages with fewer significant defects can be achieved by

growing on the substrate with smaller pore sizes. Concentration polarisation and fouling do not contribute any effect here since the pure water is used. The intrinsic hydraulic resistance is the only factor that leads to the difference. For larger substrate pores, the membrane will exhibit lower resistance and higher porosity, resulting in higher flux and permeance. As a result, there will be a high-water permeance transient at the beginning, before the structure tends toward a steady state.

4.7 Dye Rejection Performance

4.7.1 Permeance Performance for Different Dyes

Table 4.7.1 Dye Rejection Performance (* Means a New Membrane Was Used)

Dye	MW (g mol^{-1})	20nm (LMH bar $^{-1}$)	200nm (LMH bar $^{-1}$)
Orange G	452.4	5.17	103.65*
Allura Red	496.4	5.09	7.42
Congo Red	696.7	4.93	4.21
Methyl Blue	799.8	4.97	2.20
Rose Bengal	1017.6	4.89	32.14*
Average	-	5.01	67.89*

For a feed concentration of 20 ppm dye, the permeance reached steady state at approximately 8 hours after the process started. Consequently, all permeance values reported in Table 4.7.1 were measured at 24 hours to ensure steady-state conditions.

The SU-102@AAO membranes show an acceptable dye permeance. On the 20 nm substrate, permeance shows less dependence on dye molecular weight. The permeance stabilised between 4.89 and 5.17 LMH bar $^{-1}$, with an average value of 5.01 LMH bar $^{-1}$. This consistency indicates that hydraulic resistance was dominated by the selective layer, with negligible contribution from fouling or adsorption raised by dye molecules under these conditions. In contrast, permeance on the 200 nm substrate decreased with increasing dye molecular size. Larger dyes tend to accumulate and adsorb at the membrane surface. These effects form an additional resistance layer, partially blocking the more open flow pathways and lowering permeance.

Values marked with an asterisk (*) were excluded from the trend analysis because they were obtained by a new prepared membrane and under non-equilibrated adsorption conditions. For the 200 nm substrate, the newly prepared membrane was tested with Orange G and then immediately with Rose Bengal, rather than following the intended sequence

based on molecular weight. As a result, the membrane surface had not reached complete adsorption or equilibrium before the Rose Bengal measurement, leading to an extremely high permeance.

Overall, a significant decrease with increasing dye size was observed on the 200 nm substrate, but only a slight decrease was observed on the 20 nm substrate.

4.7.2 Long-term Rejection Performance for Different Dyes

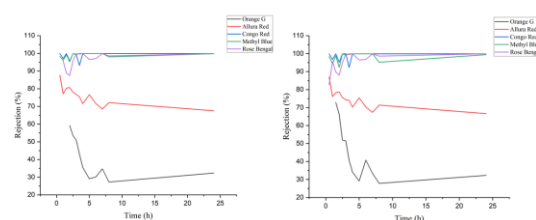


Figure 4.7.1 Long-term Rejection Performance of SU-102@AAO (20nm) Measured with Feed (left) and Retentate (right) Concentration for Different Dyes

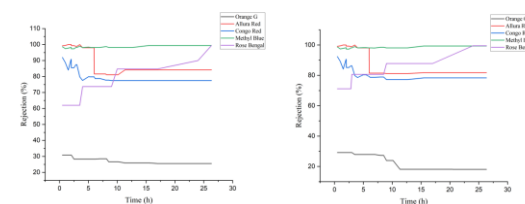


Figure 4.7.2 Long-term Rejection Performance of SU-102@AAO (200nm) Measured with Feed (left) and Retentate (right) Concentration for Different Dyes

Figure 4.7.1 shows that for the 20 nm substrate, the rejection performance of Rose Bengal, Methyl Blue, and Congo Red can approach 99.9% at around 8 hours and maintain this high rejection to the end of the experiment (24 hours). This supports the long-term stability of rejection for larger dye molecules. Also, Figure 4.7.2 shows that for the 200 nm, the rejection of Methyl Blue and Rose Bengal ultimately stabilizes at around 99.5%. This result is consistent with their large molecular sizes, which are strongly excluded by the narrow SU-102 pathways, resulting in size sieving. Moreover, this steric hindrance is further reinforced by electrostatic repulsion.

For the membrane grown on a 20 nm substrate, the Allura Red stabilizes at around 70%, and the Orange G at around 30%. For the membrane grown on a 200 nm substrate, Allura Red stabilizes at about 81%, whereas Orange G stabilizes at only about 20%. Since both dyes have similar molecular weights, the main contributor to this large difference in rejection is electrostatic exclusion. Under a neutral solvent, Allura Red's formal charge is -2, and Orange G is -1. The higher-valence dye experiences stronger electrostatic exclusion in the anionic pores and

exhibits better rejection performance. Therefore, long-term performance is governed by both the intrinsic steric hindrance and electrostatic selectivity rather than start-up holdup. The detailed descriptions of how charge and size affect will be further discussed in section 4.7.3.

Moreover, both retentate and feed concentrations are used to calculate rejection for comparison. Using retentate concentration reduces early transients, but it still shows a similar trend to feed concentration, which again shows the small effect of start-up transients on steady-state behaviour.

4.7.3 Molecular Weight Cut Off & Charge Effects under Neutral Solvent

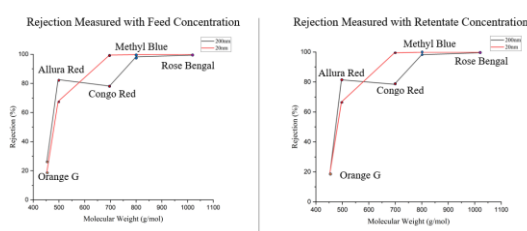


Figure 4.7.3 Steady State Rejection for Different Dyes (Red: 20nm, Black: 200nm)

For a feed dye concentration of 20 ppm and at steady state, the SU-102@AAO membrane exhibits increasing rejection performance as the dye molecular weight increases. For both performance tested by a 20 nm substrate and a 200 nm substrate, and rejection performance calculated by feed and retentate concentrations, Figure 4.7.3 shows a similar trend. This agreement indicates that the observed selectivity is intrinsic to the membrane properties and has a minimal effect from the early transient adsorption or holdup in the test cell. High-molecular-weight dyes such as Methyl Blue and Rose Bengal approach complete separation, but for lower-molecular-weight dyes, they can only partially separate. This trend is consistent with the steric hindrance mechanism.

Apart from the size-sieving effect, as discussed in section 4.7.2, the contrasting rejection behaviour of Allura Red and Orange G arises from their different formal charges despite their similar molecular sizes. To further investigate this observation, the Donnan partitioning relation provides a quantitative framework for understanding how solute valence influences transport in charged pores:

$$K = \exp\left(-\frac{zF\psi_D}{RT}\right) \quad (5)$$

where K is the equilibrium partition coefficient, z is solute valence, ψ_D is the Donnan potential set by the membrane's fixed charge, F is Faraday's constant, R is the gas constant, and T is temperature. Thus, for dyes of similar size, the more highly charged anion, which is Allura Red in this study,

experiences a larger electrostatic repulsion effect and shows higher rejection.

However, when comparing to SU-102 membranes grown on 20 nm substrate and 200 nm substrate, there's a big difference in trend for Allura Red and Congo Red. This is due to the spatial regime. When SU-102 is grown on the 20 nm AAO substrate, the selective layer more effectively bridges pores and leaves only narrow intercrystalline passages. For Allura Red and Congo Red, both have the same formal charge of -2, and therefore, they experience similar electrostatic repulsion, and the steric sieving effect dominates. As a result, since the Congo Red is larger, it rejects better than Allura Red.

However, on the 200 nm substrate, SU-102 growth is more likely to leave wider intercrystalline gaps that hydraulically couple to the large substrate pores. In this spatially screened regime, differences in effective charge density and charge distribution become important. Although Allura Red and Congo Red both have a valence of 2, the Allura Red molecule is a smaller and more compact molecule with closely spaced sulfonate groups. Its structure provides a higher effective charge density and stronger near-wall repulsion. However, Congo Red is larger and more extended, with distributed charges that are more readily counter-ion screened. Hence, it can exhibit stronger π - π interactions with the aromatic pore walls of the membrane. Under these conditions, due to the weaker effective electrostatic exclusion and greater dynamic adsorption effect, Congo Red shows a lower rejection performance in wider passages on the 200 nm substrate.

Taken together, the data show that SU-102 exhibits a great size-sieving selectivity, which is enhanced by electrostatic repulsion. It balances both steric and electrostatic effect by changing the geometry of transport pathways by controlling the substrate pore size. For a narrower passage in a 20 nm substrate, the size effect dominates, and for a wider pathway in a 200 nm substrate, the steric effect is reduced, and the effective electrostatic exclusion dominates. This analysis shows that the molecular-weight cut-off depends on the pore size of the substrate. In order to optimise the separation performance for specific pollutants that need to be remediated, the membrane can be tuned by controlling the substrate pore based on the size and charge of pollutants.

5. Discussion

For future studies, there are still several factors that need to be considered, such as substrates with different pore sizes, different types of substrates, variable dyes with different positive or negative charges and different molecular weights, different proportions in solvent composition, solvent pH

values, membrane areas, feed concentrations, and operating pressures. All of them are likely to influence the balance between rejection, permeance, and energy consumption. After collecting experimental data and conducting simulations, the systematic optimisation of these parameters would help to identify the best operating conditions to maximise separation efficiency.

For the analytical setup, there are some limitations of the instrument that cause the unreliability of the data. For the discussion of the dye charge part in this study relies on the theoretical molecular structures, which are not reliable under different conditions. Using zeta-potential measurement is more accurate, but the instrument in the Analytical Lab at Imperial College was under maintenance. Once the instrument is repaired, direct determination of dye charge would provide a more detailed and quantitative evidence on how membrane electrostatically interact with solute. Hence, a more in-depth interpretation of the observed charge-governed selectivity can be investigated.

Moreover, the NanoDrop UV-Vis instrument was used in this study to measure the absorbance, but it showed substantial scatter between replicate scans. This variability indicates that other, more precise spectrophotometers, such as UV-Vis MT or UV-Vis Shimadzu, should be applied. These instruments would enhance the accuracy and reproducibility

For size sieving behaviour, there are still several improvements that can be applied. Since both the 20 nm and 200 nm substrate membranes showed anomalous permeance under certain dye conditions, sequential testing of dyes in increasing molecular weight would minimise the effect of non-equilibrated fouling. Also, more substrates with different pore sizes should be used. As a result, it can construct a more reliable rejection and molecular-weight calibration curve.

Theoretically, the retentate concentrations did not exhibit the expected increasing trend. This leads to a more complicated comprehension of the transition from adsorption-dominated behaviour to steady-state nanofiltration. As a result, extending filtration duration and investigating the origin of deviation can clarify adsorption dynamics, mass-transfer limitations, and the starting point of a true steady-state separation.

Overall, once adopting all limitations and improvements, the steric and electrostatic selectivity of SU-102@AAO in this study can be more accurate and precise.

6. Conclusion

In this study, an in-situ growth of composite SU-102 membrane into Anodic Aluminium Oxide (AAO) substrate and its characteristics, permeation, and rejection performance, tested with different negatively charged dyes, are reported. The in-situ growth of SU-102@AAO membrane was achieved by immersing the substrate in the precursor solution prepared. The pore size and density are controlled by the selection of different substrates, which are 20 nm and 200 nm in this report. SEM was employed for the morphology, and XRD and FTIR were used to confirm the successful formation of the metal-organic framework. The Contact Angle test is used to test the hydrophilicity, and the TGA Analysis is used to measure the decomposition of the membrane under different temperature regions. SU-102@AAO membrane grown on a 20 nm substrate shows a high and stable water permeance of 410 LMH bar⁻¹ and membrane grown on a 200 nm substrate shows a transient initially and much higher water permeance at steady state, about 780 LMH bar⁻¹. For the dye rejection performance, the membrane selectivity combines both the steric size sieving and the electrostatic repulsion effect. Such performance results exhibit the excellent potential of the SU-102@AAO membrane in the water treatment area due to its tunable pore size and near-complete rejection for negatively charged dyes, which have a large molecular size.

7. Acknowledgements

The authors gratefully acknowledge the invaluable academic guidance and continuous support provided by Ms. Yuting Wu and Ms. Mengjiao Zhai throughout this research project.

Furthermore, the authors are also very thankful for the technical support and analytical services provided by the Analytical Laboratory within the Department of Chemical Engineering at Imperial College London. Special thanks are due to analytical technician Mr. Koho Cheung for his expert assistance in the Analytical Laboratory.

8. References

- [1] Motshekga, S.C., Oyewo, O.A. & Makgato, S.S. (2024). Recent and prospects of synthesis and application of metal-organic frameworks (MOFs) in water treatment: a review. *Journal of Inorganic and Organometallic Polymers and Materials*, 34, 3907–3930.
<https://doi.org/10.1007/s10904-024-03063-x>
- [2] Dutta, S., Adhikary, S., Bhattacharya, S., Roy, D., Chatterjee, S., Chakraborty, A., Banerjee, D., Ganguly, A., Nanda, S. & Rajak, P. (2024). Contamination of textile dyes in aquatic environment: Adverse impacts on aquatic ecosystem

and human health, and its management using bioremediation. *Journal of Environmental Management*, 353, 120103.

<https://doi.org/10.1016/j.jenvman.2024.120103>

[3] Hoque, M.B., Oyshi, T.H., Hannan, M.A., Haque, P., Rahman, M.M., Shahid, M.A. & Sheikh, S. (2024). Unraveling the ecological footprint of textile dyes: A growing environmental concern. *Pollution Study*, 5(2), 3014. <https://doi.org/10.54517/ps.v5i2.3014>

[4] Chai, J.S., Jung, B., Kim, H.-H., Heo, J., Park, C.M., Jang, M., Nam, S.-N., Huang, Y., Jun, B.-M. & Yoon, Y. (2024) A review of metal–organic framework-based membranes for the removal of emerging contaminants from water. *Journal of Water Process Engineering*, 68, 106456. <https://doi.org/10.1016/j.jwpe.2024.106456>

[5] Moghadam, F. & Park, H.B. (2018). Two-dimensional materials: an emerging platform for gas separation membranes. *Current Opinion in Chemical Engineering*, 20, 28–38. <https://doi.org/10.1016/j.coche.2018.02.004>

[6] Chen, Z., Hanna, S.L., Redfern, L.R., Alezi, D., Islamoglu, T. & Farha, O.K. (2019). Reticular chemistry in the rational synthesis of functional zirconium cluster-based MOFs. *Coordination Chemistry Reviews*, 386, 32–49. <https://doi.org/10.1016/j.ccr.2019.01.017>

[7] Xia, J., Gao, Y. & Yu, G. (2021). Tetracycline removal from aqueous solution using zirconium-based metal–organic frameworks (Zr-MOFs) with different pore size and topology: Adsorption isotherm, kinetic and mechanism studies. *Journal of Colloid and Interface Science*, 590, 495–505. <https://doi.org/10.1016/j.jcis.2021.01.046>

[8] Clark, C.A., Heck, K.N., Powell, C.D. & Wong, M.S. (2019). Highly defective UiO-66 materials for the adsorptive removal of perfluorooctanesulfonate. *ACS Sustainable Chemistry & Engineering*, 7(7), 6619–6628. <https://doi.org/10.1021/acssuschemeng.8b05572>

[9] Huang, L., He, M., Chen, B. & Hu, B. (2018). Magnetic Zr-MOFs nanocomposites for rapid removal of heavy metal ions and dyes from water. *Chemosphere*, 199, 435–444. <https://doi.org/10.1016/j.chemosphere.2018.02.019>

[10] Azhar, M.R., Abid, H.R., Sun, H., Periasamy, V., Tadé, M.O. & Wang, S. (2017). One-pot synthesis of binary metal organic frameworks (HKUST-1 and UiO-66) for enhanced adsorptive removal of water contaminants. *Journal of Colloid*

and Interface Science, 490, 685–694. <https://doi.org/10.1016/j.jcis.2016.11.100>

[11] Aslam, S., Zeng, J., Subhan, F., Li, M., Lyu, F., Li, Y. & Yan, Z. (2017). In situ one-step synthesis of Fe₃O₄@MIL-100(Fe) core–shells for adsorption of methylene blue from water. *Journal of Colloid and Interface Science*, 505, 186–195. <https://doi.org/10.1016/j.jcis.2017.05.090>

[12] Wu, J., Zhou, J., Zhang, S., Alsaedi, A., Hayat, T., Li, J. & Song, Y. (2019). Efficient removal of metal contaminants by EDTA modified MOF from aqueous solutions. *Journal of Colloid and Interface Science*, 555, 403–412. <https://doi.org/10.1016/j.jcis.2019.07.108>

[13] Wang, C., Lin, G., Zhao, J., Wang, S., Zhang, L., Xi, Y., Li, X. & Ying, Y. (2020). Highly selective recovery of Au(III) from wastewater by thioctic acid modified Zr-MOF: Experiment and DFT calculation. *Chemical Engineering Journal*, 380, 122511. <https://doi.org/10.1016/j.cej.2019.122511>

[14] Tripathi, S., Sreenivasulu, B., Suresh, A., Brahmmananda Rao, C.V.S. & Sivaraman, N. (2020). Assorted functionality-appended UiO-66-NH₂ for highly efficient uranium(VI) sorption at acidic/neutral/basic pH. *RSC Advances*, 10, 14650–14661. <https://doi.org/10.1039/D0RA00410C>

[15] Abney, C.W., Taylor-Pashow, K.M.L., Russell, S.R., Chen, Y., Samantaray, R., Lockard, J.V. & Lin, W. (2014). Topotactic transformations of metal–organic frameworks to highly porous and stable inorganic sorbents for efficient radionuclide sequestration. *Chemistry of Materials*, 26(18), 5231–5243. <https://doi.org/10.1021/cm501894h>

[16] Wang, Y., Yan, B., Liu, J., Wang, R., Rao, P. & Liu, Y. (2025). Advances in MOF membrane strategies for selective lithium extraction from salt lake brine. *Advanced Membranes*, 5, 100156. <https://doi.org/10.1016/j.advmem.2025.100156>

[17] Wu, M., Sun, Y., Ji, T., Yu, K., Liu, L., He, Y., Yan, J., Meng, S., Hu, W., Fan, X., Du, D. & Liu, Y., 2023. Fabrication of water-stable MOF-808 membrane for efficient salt/dye separation. *Journal of Membrane Science*, 686, 122023. <https://doi.org/10.1016/j.memsci.2023.122023u>

[18] Fang, S.-Y., Zhang, P., Gong, J.-L., Tang, L., Zeng, G.-M., Song, B., Cao, W.-C., Li, J. & Ye, J. (2020) Construction of highly water-stable metal–organic framework UiO-66 thin-film composite membrane for dyes and antibiotics separation.

Chemical Engineering Journal, 385, 123400.
<https://doi.org/10.1016/j.ccej.2019.123400>

[19] Grape, E.S., Chacón-García, A.J., Rojas, S., Pérez, Y., Jaworski, A., Nero, M., Åhlén, M., Martínez-Ahumada, E., Galetsa Feindt, A.E., Pepillo, M., Narongin-Fujikawa, M., Ibarra, I.A., Cheung, O., Baresel, C., Willhammar, T., Horcajada, P. & Inge, A.K. (2023) 'Removal of pharmaceutical pollutants from effluent by a plant-based metal–organic framework', *Nature Water*, 1, pp. 433–442.
<https://doi.org/10.1038/s44221-023-00077-5>

[20] Reinsch, H. (2016). "Green" synthesis of metal–organic frameworks. *European Journal of Inorganic Chemistry*, 2016(27), 4490–4500.
<https://doi.org/10.1002/ejic.201600286>

[21] Zhao, S.-N., Zhang, Y., Song, S.-Y. & Zhang, H.-J. (2019). Design strategies and applications of charged metal–organic frameworks. *Coordination Chemistry Reviews*, 398, 113007.
<https://doi.org/10.1016/j.ccr.2019.07.004>

[22] Canivet, J., Fateeva, A., Guo, Y., Coasne, B. & Farrusseng, D. (2014). Water adsorption in metal–organic frameworks: fundamentals and applications. *Chemical Society Reviews*, 43(16), 5594–5617.
<https://doi.org/10.1039/C4CS00078A>

[23] Lee, H.J., Shirke, Y.M., Kim, J., Yu, H.J., Yoo, C.H., Back, S., Jeon, J.-D. & Lee, J.S. (2023). Tailoring molecular structures of UiO-66-NH₂ for high-performance H₂O/N₂ separation membranes: A synergistic effect of hydrophilic modification and defect engineering. *Journal of Membrane Science*, 665, 121096.
<https://doi.org/10.1016/j.memsci.2022.121096>

[24] Yu, Y., Wu, X.-J., Zhao, M., Ma, Q., Chen, J., Chen, B., Sindoro, M., Yang, J., Han, S., Lu, Q. & Zhang, H. (2017). Anodized aluminum oxide templated synthesis of metal–organic frameworks used as membrane reactors. *Angewandte Chemie International Edition*, 56(2), 578–581.
<https://doi.org/10.1002/anie.201610291>

[25] Clark, J. (n.d.). The Beer-Lambert Law. *Chemistry LibreTexts*. Available at: <https://chem.libretexts.org/>

[26] National Center for Biotechnology Information (NCBI). (n.d.). *PubChem*. National Library of Medicine. Available at: <https://pubchem.ncbi.nlm.nih.gov/>

Techno-Economic Analysis of Crystallisation for the Continuous Production of Paracetamol with Solvent Recycling

Shaurya Choithani and Aryan Salastekar

Department of Chemical Engineering, Imperial College London, U.K.

1 Abstract

This study addresses the high solvent intensity of pharmaceutical manufacturing by developing an integrated Computer-Aided Molecular and Process Design (CAMPD) framework for the continuous cooling crystallization of paracetamol. A rigorous steady-state model was constructed in gPROMS to simulate a closed-loop flowsheet featuring solvent recycling and the rejection of synthesis impurities (Metacetamol and Acetanilide). The process physics were resolved using the SAFT-gamma-Mie equation of state, which provided physically consistent predictions for phase behaviour, enthalpy, and solubility. The simulation was scaled to an industrial capacity of **187.5 kg·h⁻¹** (approximately 1% of global demand) and subjected to a detailed Techno-Economic Analysis (TEA). Capital costs were estimated via a Chilton-style cascade, while operating expenditures integrated solvent, utility, and labour charges. The baseline process is projected to be profitable from the first year of operation; assuming a selling price of **£5.23 kg⁻¹**, the facility yields a projected lifetime revenue of **£134 million** over a 20-year horizon. Cost breakdown analysis reveals that crystallizer volume is the dominant driver of Capital Expenditure (CAPEX), whereas solvent make-up rates govern Operating Expenditure (OPEX). Sensitivity analysis demonstrates that while increasing the evaporation fraction enhances solvent recovery, it leads to disproportionately higher thermal duties, resulting in operating costs more than double those of the Base Case. The study concludes that the unmodified Base Case offers the optimal techno-economic configuration, balancing low operating costs with competitive Process Mass Intensity (PMI) and E-factor metrics without requiring aggressive intensification.

2 Introduction

Crystallisation is the keystone process of small-molecule pharmaceutical manufacturing, with an estimated **90%** of products handled in crystalline form during production [1]. Yet, despite its utility as both a purification and isolation step, it is also among the most solvent-intensive operations. Solvents often account for over **80%** of the total Process Mass Intensity (PMI) during manufacturing [2]. Traditional batch processes often discharge these spent solvents rather than recover them [3], resulting in wasteful practices where E-factors frequently range from **25** to **100 kg** of waste per kg of product [2].

Continuous manufacturing offers a compelling alternative, particularly for crystallization, where integration with solvent recovery is more feasible. End-to-end continuous processes have demonstrated lower environmental impact, achieving footprint reductions of up to **90%** and operating cost savings of **30-50%** [4] while maintaining product quality [1]. As a result, solvent recycling in continuous crystallization now presents a practical path to reducing solvent use, energy input, and emissions [5], aligning with industry goals for greener manufacturing [4].

This study develops and evaluates a continuous, recycle-enabled cooling crystallisation flowsheet for paracetamol in methanol. Two realistic synthesis impurities (metacetamol and acetanilide) are included to ensure the separation performance reflects industrial complexity [3]. The objectives are to quantify solvent reduction and economic impact while maintaining target purity, and to map the trade-offs between recycle ratio,

yield, and energy use. Metrics include Process Mass Intensity [2], energy duty per kilogram of API, and a consistent capital versus operating cost perspective [4].

The approach combines first-principles modelling and standard techno-economic scaling. A flowsheet with MSMR crystallisers, agitated Nutsche filtration and washing, a flash evaporator, a condenser, a recycle tank, and a process-side heat exchanger is implemented in gPROMS. Thermophysical properties and phase behaviour are predicted using SAFT- γ -Mie through a foreign-object interface [6]. Steady-state mass and energy balances drive equipment sizing from volumetric flows and residence times, and costs are estimated using a Chilton-style cascade scaled with CEPCI.

3 Background

3.1 Project Objective

The primary objective of this project is to apply Computer-Aided Molecular and Process Design (CAMPD) methods to simulate a crystallization flowsheet with fully integrated solvent and API recycling. Within this framework, energy loads, utility usage, and process costs are calculated simultaneously alongside the mass balance. The process configuration is scaled to meet a production target of **1%** of global paracetamol demand, defined as **187.5 kg/hr** of API over **8,000** operating hours per year. Finally, the study conducts a Techno-Economic Analysis (TEA) to quantify the economic viability and operational longevity of this continuous processing approach.

3.2 Justifications

The paracetamol-methanol system was selected for its industrial relevance, favourable thermodynamic properties, and extensive characterization in the literature [7,8]. Methanol exhibits a steep temperature-dependent solubility gradient for paracetamol, enabling high crystallization yields at modest sub-ambient temperatures [7,8].

Furthermore, methanol's low normal boiling point (64.7 °C) and high volatility significantly reduce the energy duty required for solvent recovery via flash evaporation compared to higher-boiling alternatives like isopropanol [9]. The solvent is also highly compatible with the SAFT- γ -Mie equation of state, ensuring that the liquid enthalpy, density, and VLE predictions required for equipment sizing are robust [6].

4 Methodology and Process Description

The simulation framework for this study was developed through a four-stage hierarchical approach. First, a baseline material balance was established at the laboratory scale to define the fundamental mass flows. Second, the topology was expanded to include a closed-loop solvent recovery system, integrating secondary crystallization and filtration units to maximize yield.

4.1 Process Architecture and Operating Conditions

The continuous manufacturing line (Fig 4.2) begins with a dissolution stage where paracetamol and realistic synthesis impurities (acetanilide (precursor) and metacetamol (Regio-isomer)) are dissolved in methanol at 43.85 °C. To reflect industrial complexity, impurities are modelled at a fixed mass ratio of 1/30th of the API in the feed. This solution feeds Crystallizer 1, a jacketed Mixed Suspension Mixed Product Removal (MSMPR)

Following the mixing stage, the process stream enters the solvent recovery loop. To separate the solvent from non-volatile impurities, the stream is flashed in an evaporator operating at 69.57 °C, which corresponds to the calculated bubble point of the $\approx 95\%$ methanol mixture. The methanol vapor is recovered overhead and condensed at 20 °C, while the impurity-rich liquid bottoms are directed to a secondary recovery stage. This secondary section consists of Crystallizer 2 (operating at

To reflect realistic industrial separation challenges, two specific impurities were tracked: acetanilide (the synthesis precursor) and Metacetamol (Regio-isomer) [10]. Both structurally resemble the API and tend to remain in the mother liquor under the chosen operating conditions. Including these species rigorously tests the efficiency of the filtration, washing, and purge loops without introducing unnecessary computational overhead, ensuring the findings remain applicable to typical pharmaceutical crystallization processes (Fig 4.1)

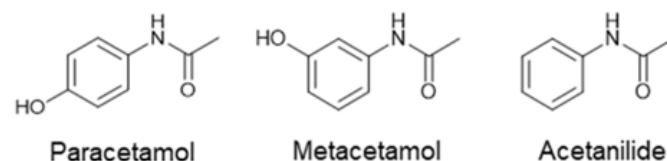


Figure 4.1: Chemical Structures of Process Compounds

Third, the SAFT- γ -Mie equation of state was embedded into the model to resolve the energy balances. Finally, the model was scaled up to an industrial capacity of 187.5 kg/hr (99.8% purity) and converged to a stable steady state, representing an annual production of approximately 1% of global paracetamol demand over an 8,000-hour operating year.

unit operating at 5 °C. This sub-ambient temperature generates the specific supersaturation required for the primary product yield. The resulting slurry is processed in Filter 1, where the filter cake is washed with fresh solvent. The mother liquor from the primary filter, now at 5 °C is combined with the wash filtrate and the recycle stream in a mixing unit. The resulting temperature of this combined stream is calculated at 8.29 °C based on the specific enthalpies and mass ratios of the inflowing streams.

20 °C) and Filter 2, designed to capture residual API that would otherwise be lost to the purge. A controlled purge stream is extracted from the secondary filtrate to prevent the accumulation of impurities before the recovered solvent is returned to the feed tank. A process-side heat exchanger on the recycle line ensures the solvent returns to the dissolution vessel at the required 43.85 °C, closing the thermal loop.

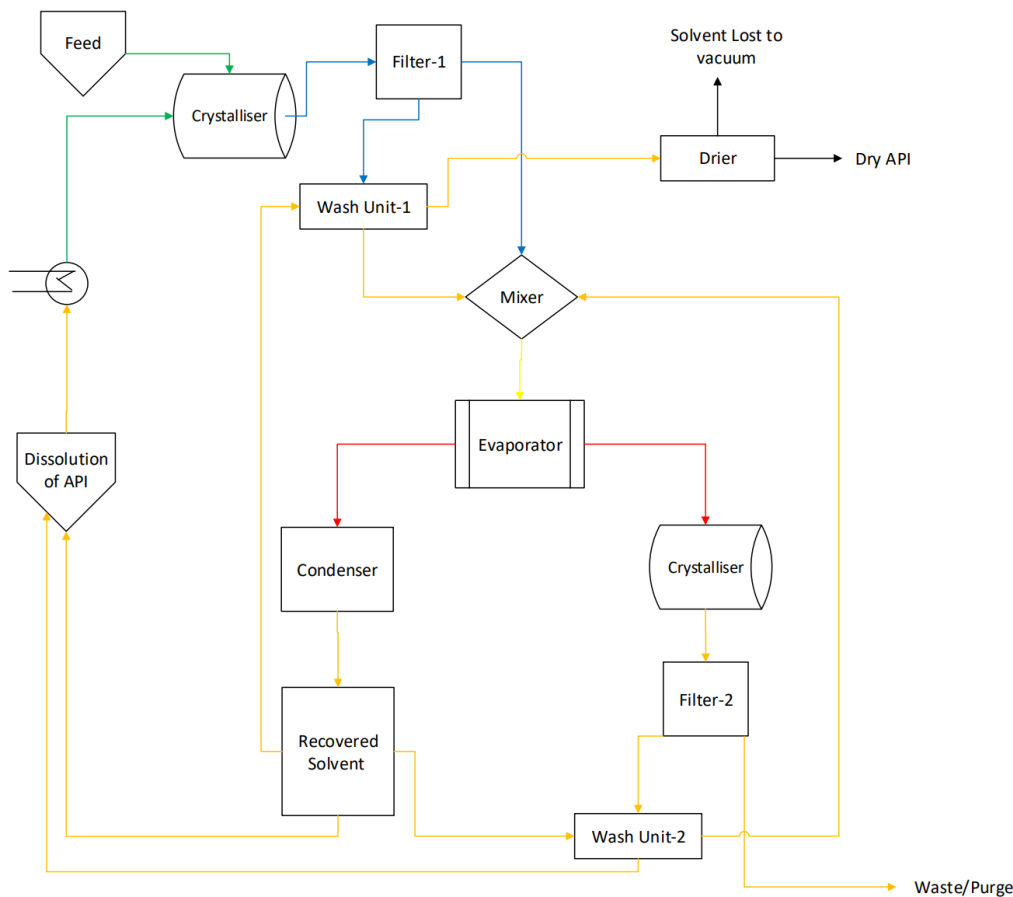


Figure 4.2: Process Diagram

4.2 Computational Framework (gPROMS & SAFT- γ -Mie)

The process physics are resolved using a rigorous steady-state model in gPROMS. The SAFT- γ -Mie equation of state is coupled via a foreign object interface to provide physically consistent predictions for the fluid phases. The process model utilizes a hybrid approach to property characterization. The foreign object interface executes dynamic calls for LiquidEnthalpy and VaporEnthalpy, ensuring accurate energy balances for the fluid streams. Volumetric sizing relies on LiquidDensity calculations, while the flash evaporator utilizes the PVFMolFlash call to determine the vapor-liquid split, flash temperature, and enthalpy of vaporization. Additionally, the specific heat capacity (C_p) of the liquid API is derived directly from the equation of state.

However, due to limitations in the SAFT property package regarding solid-phase characterization, specific solid-state parameters were sourced from literature. These include the Enthalpy of Fusion and the Specific Heat Capacity (C_p) of the solid API. Similarly, the specific heat capacity of water used in the utility cooling jackets and Heat Exchanger is treated as a fixed parameter.

Table 4.2: Fixed Thermodynamic Parameters sourced from Literature

Parameter	Value	Unit	Source
Enthalpy of Fusion (API)	27.0	kJ/mol	[11]
Solid Specific Heat (API)	1.60	kJ/(kg·K)	[11]
Specific Heat (Water)	4.18	kJ/(kg·K)	[12]

4.3 Mathematical Formulation

The solubility of paracetamol in methanol (w_s), a critical parameter for yield prediction, is defined by an empirical temperature-dependent exponential regression derived from literature data [7]:

$$w_s(T) = 0.1929 e^{0.0217 T [^{\circ}\text{C}]}$$

The energy requirements for the crystallizers are computed by integrating the specific heat capacities (sourced from both SAFT and literature) and phase change enthalpies. The total cooling duty (Q_C^-) is calculated by summing the sensible heat change of the

liquid phase (from inlet T_9 to outlet T_7) and the total enthalpy change of the crystallizing API. This includes the sensible heat of the liquid API (SAFT-derived), the heat of mixing (ΔH_m), and the sensible heat of the solid crystal lattice (Literature-derived) [13]:

$$Q_C^- = \sum_{i \in C} N_{i,9} \int_{T_9}^{T_7} C_{Pi}^{(L)} dT + N_{D,10} \left(\int_{T_9}^{T_7} C_{PAPI}^{(L)} dT + \Delta H_m(T_m) + \int_{T_9}^{T_m} C_{PAPI}^{(S)} dT \right)$$

(further duty equations for the other units are available in the appendix)

5 Equipment Sizing and Economic Evaluation

5.1 Capital Expenditure (CAPEX) Assessment

The economic assessment of the proposed continuous manufacturing process relies on the Chilton method, a classical and widely adopted approach for estimating capital outlay at the early design stage. This technique utilizes empirical cost correlations derived from historical industrial data, which are updated to present-day values using the Chemical Engineering Plant Cost Index (CEPCI). In this study, the Chilton method is integrated directly into the gPROMS model to estimate the Total Plant Production Cost (TPPC) and Total Capital Expenditure (CAPEX). The valuation is performed in two distinct steps: first, determining the physical capacity of each unit based on the simulated process performance, and second, converting that capacity into a monetary cost using capacity-scaling laws (A.6).

The foundational calculation for every major unit operation (including crystallizers, filters, heat exchangers, and vessels) is the determination of the Base Purchase Cost (P_B). This is calculated using a power-law scaling equation where the reference cost is adjusted by the ratio of the actual required capacity (S_B) to a reference capacity (S_A), raised to an economy-of-scale exponent (n) with a design factor for the materials of construction (f):

$$P_B = P_A \left(\frac{S_B}{S_A} \right)^n (1 + f)$$

Once the Free-on-Board (FOB) purchase cost is established, the Chilton method applies a sequential cascade of factors to estimate the fully installed cost. First, an installation factor of 1.43 is applied to the FOB cost to derive the Installed Equipment Cost (IEC), which accounts for labour, foundations, and ancillary connections. Subsequently, a piping and instrumentation factor of 0.42 is applied to the IEC to form the Total Plant Production Cost (TPPC)[5]:

$$\begin{aligned} WC &= 0.035 MAT_{Annual} \\ IEC &= 1.43 FOB \\ PPI &= 0.42 IEC \\ TPPC &= IEC + PPI \end{aligned}$$

The cost is finally escalated by a factor of 1.30 to determine the Bare Module Cost (BLIC), representing construction overheads, field expenses, and contractor fees. The final CAPEX estimate is derived by summing the BLIC across all units and incorporating additional allowances for contingency and working capital [5].

$$\begin{aligned} CC &= 0.2 BLIC \\ WCC &= WC + CC \\ BLIC &= 1.3 TPPC \\ CAPEX &= BLIC + WCC \end{aligned}$$

To utilize these cost equations, the required capacity (S_B) for each unit is determined dynamically within the simulation. The gPROMS model calculates this by strictly coupling the steady-state mass balance with thermodynamic properties predicted by SAFT- γ -Mie. Specifically, the mass flow rates ($\dot{m}_i$) derived from the mass balance are divided by the stream-specific densities ($\rho_{mixture}$) computed via LiquidDensity calls to yield accurate volumetric flow rates ($\dot{V}$):

$$\dot{V} = \frac{\sum_i \dot{m}_i}{\rho_{mixture}}$$

The sizing logic for specific unit operations reflects standard pharmaceutical engineering practices. The Mixed Suspension Mixed Product Removal (MSMPR) crystallizers are sized based on the volumetric outflow of the slurry. To ensure sufficient crystal growth, the total vessel volume (S_B) is determined by multiplying the slurry flow rate by a residence time (τ) of **3 hours** [14]:

$$S_B = \dot{V} \times \tau$$

Similarly, the Agitated Nutsche Filter-Driers (ANFD) are designed based on the volumetric flow of the wet cake. The cake density is predicted using liquid correlations and a fixed porosity parameter, while the required filtration area (A_{filter}) is calculated by applying a standard superficial filtration velocity (v_{sup}) of **0.5 m/h** [12]:

$$A_{filter} = \frac{\dot{V}_{cake}}{v_{sup}}$$

Downstream processing units are sized to match the throughput of the purification train. The recycle heat exchanger is sized according to the thermal duty (Q_{HX}^+) required to return the recycle stream to the feed temperature. The model computes the Log-Mean Temperature Difference (LMTD) and utilizes a global heat transfer coefficient (U) to solve for the required heat transfer area (A_{HX}):

$$A_{HX} = \frac{Q_{HX}^+}{U \cdot LMTD}$$

Finally, the solvent recovery section consists of a flash evaporator and a condenser receiver. These vessels are sized as vertical drums based on the liquid hold-up of the bottoms and condensate streams, respectively. A short residence time of **5 minutes** (0.083 hours) is

applied to these units to minimize the inventory of hot solvent, yielding the vessel volume (V_{vessel})[14]:

$$V_{vessel} = \dot{V} \times \tau$$

5.2 Operating Expenditure (OPEX) Assessment

The Operating Expenditure (OPEX) constitutes the recurring financial burden of the manufacturing process and captures the trade-offs between the solvent savings achieved by the recycle loop and the energy intensity required to sustain it. The analysis aggregates variable

costs, such as raw materials and utilities, with fixed production costs into a single annualized metric. The total cost per unit of product is governed by the following factorial equation:

$$OPEX_{E/kgAPI} = \frac{C_{MAT} + f_{maint}BLIC + f_{lab}BLIC + f_{oh}f_{lab}BLIC + (Q_{heat}c_{heat} + Q_{cool}c_{cool})t_{op}}{m_{API}}$$

Variable Costs: Materials and Utilities

Material expenditure (C_{MAT}) is driven by the consumption of fresh feedstocks and the makeup solvent required to replenish process losses. While the recycle loop significantly reduces steady-state demand, solvent losses through the vacuum drier, filtration cake hold-up, and impurity purge must be continuously replaced. The cost analysis utilizes a representative bulk price of **£0.40 kg⁻¹** for industrial-grade methanol [15,16]. Paracetamol feed costs are derived internally from the upstream synthesis model rather than treated as a variable input in this specific calculation.

Utility costs reflect the thermodynamic duties computed via the SAFT- γ -Mie energy balances. The total energy expenditure is derived by converting the thermal loads of the evaporator (Q_{heat}) and crystallizers (Q_{cool}) into annual consumption values based on an operating year (t_{op}) of **8,000 hours**. These loads are then multiplied by specific utility tariffs. To represent a realistic UK industrial context, the model applies rates of **£12.00**

GJ⁻¹ for low-pressure steam [17] and **£0.35 GJ⁻¹** for cooling water [18], alongside an electricity tariff of **£0.12 kWh⁻¹** [19]. Waste disposal costs are accounted for at **£0.50 kg⁻¹** for spent solvent [20] and **£0.20 kg⁻¹** for solid waste [21] (A.7)

Fixed Production Costs

Fixed costs typically comprise labour, maintenance, and plant overheads. These are estimated using a factorial approach based on the Bare Module Cost ($BLIC$) and operating requirements. Maintenance costs are calculated as **3%** of the fixed capital investment ($f_{maint} = 0.03$) [17] to represent routine servicing and minor repairs. Operating labour is modelled as a function of the plant scale and capital complexity ($f_{lab}BLIC$) and is anchored by a standard operator salary of **£45,000 yr⁻¹** [22]. Finally, plant overheads cover supervision, administration, and general facilities. These are computed as **60%** of the operating labour cost ($f_{oh} = 0.60$) [16]. The sum of these components is normalized by the annual API production (m_{API}) to yield the final economic performance indicator.

6 Results and Discussion

Based on the base case of the model, the following results for capital and operating expenditure were obtained from the Techno-Economic Analysis [16][17]

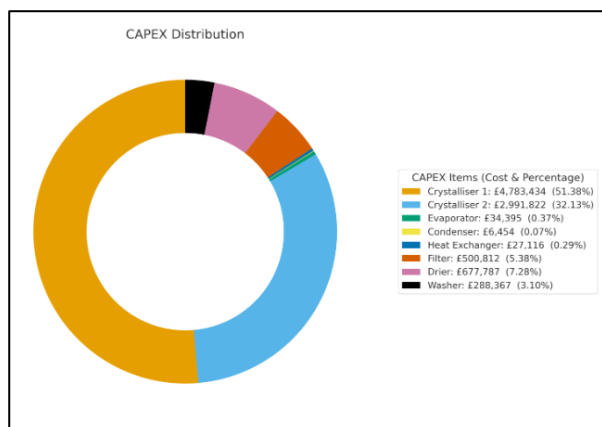


Figure 6.1: CAPEX by Component for Base Case

The key learnings from the CAPEX graph are (Figure 6.1):

- (1) The capital cost breakdown shows that the Crystallizers dominate CAPEX, with Crystallizer 1 contributing around half of the total cost and Crystallizer 2 contributing roughly a third. This reflects their large working volumes, which are required to meet residence time and throughput requirements.
- (2) Other units such as the filter, dryer, and washer contribute modestly, while the evaporator, condenser, and heat exchanger are negligible in comparison.

The operating expenditure graph tells us that (Figure 6.2):

- (1) Material costs dominate OPEX at just over half, largely driven by solvent make-up and raw material usage.
- (2) Labor accounts for around a quarter, reflecting the scale of installed equipment.
- (3) Overheads and maintenance make smaller but non-negligible contributions.

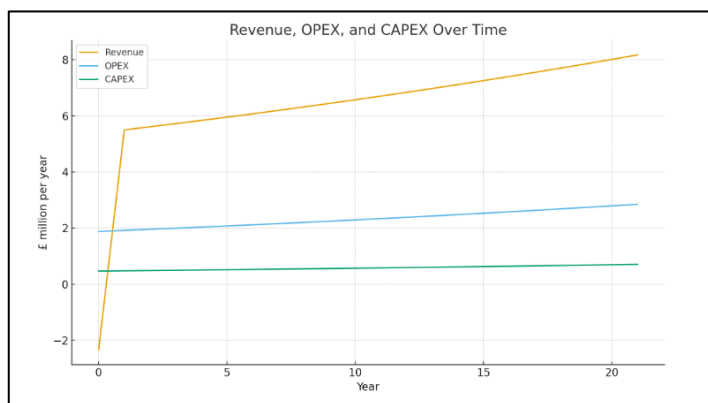


Figure 6.1.1: Base-case annual revenue and operating expenditure over the 20-year project period. Revenue for each year depicted on a single-year basis (revenues from previous years are not added to the figure of a given year).

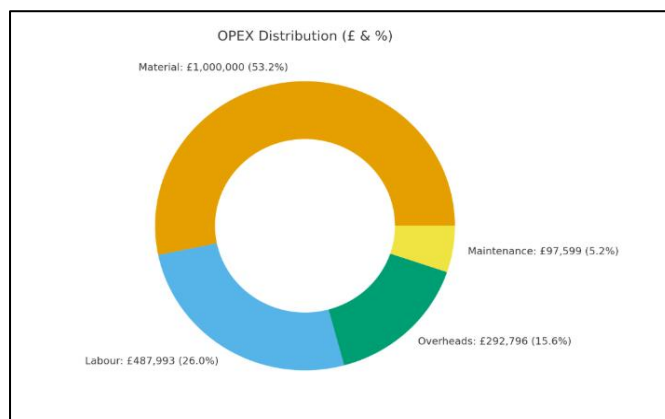


Figure 6.2: OPEX for Base Case

Overall, the base case highlights that crystallization governs capital intensity, while solvent usage drives operating cost, providing a clear benchmark for evaluating alternative design cases.

6.1 Revenue

Revenue (Figure 6.1.1) is estimated from the annual mass of API produced and the unit selling price [23]. The process operates at **187.5 kg·h⁻¹** for **8000 hours per year**, giving an annual throughput of **1.5 million kg**. Using a selling price of **£5.23 kg⁻¹**, the Year-1 revenue is **£5.5 million** [23]. Revenue in the start-up year is assumed to be zero. A **2% annual inflation rate** is applied over the 20-year project horizon to account for long-term economic conditions. Inflation is compounded yearly for revenue, operating expenditure, and capital expenditure [16][18], resulting in an inflation-adjusted financial profile appropriate for long-term techno-economic analysis.

The lifetime revenue over a 20-year project basis is projected to be £134 million (obtained by summing revenue from all years).



Figure 6.1.2: Sensitivity analysis of paracetamol selling price, showing revenue variation under $\pm 10\%$ and $\pm 20\%$ price scenarios over 20 years. Revenue from each year is depicted on a single-year basis (revenues are not added to the figure of a given year).

6.2 Sensitivity to Selling Price of Paracetamol

Revenue (Figure 6.1.2) grows steadily with inflation across all scenarios. Higher prices shift the revenue trajectory upward almost proportionally, while price reductions compress it. By Year 20, a +20% price uplift generates roughly £1.6 million more revenue per year than the base case, whereas a -20% cut removes about the same magnitude. This graph is indicative of how sensitive the profitability of the project based on the selling price of paracetamol [15][16]

6.3 Creating Different Cases

A key aim of the project was to evaluate multiple design configurations and identify the most promising process option. Four critical design variables were selected based on their strong influence on solvent recovery,

process efficiency, and economic outcomes—particularly product yield, energy consumption, and operating cost in crystallisation [24][26][27]. These variables were:

1. Evaporation fraction (α_{evap}): 0.50 and 0.80
2. Wash solvent mass fraction: 0.30 and 0.70
3. Solvent hold-up in the cake: 0.20 and 0.35 kg MeOH per kg API
4. Filter superficial velocity (TEA_FI_v_sup_mph): 0.30 and 0.80 $\text{m}\cdot\text{h}^{-1}$

Each variable was varied between realistic upper and lower bounds informed by industrial practice [25]. Holding three variables constant while varying one produced 16 possible combinations across the design space. From these, eight representative cases were selected for detailed simulation (Table 6.3.1), ensuring all configurations correspond to feasible and implementable operating conditions in pharmaceutical processes.

Table 6.3.1: Cases for Key Design Variables

Case	Evaporation Fraction, α_{evap}	Wash Mass Fraction (kg wash/kg API)	Cake Solvent Hold-up (kg MeOH/kg API)	Filter Superficial Velocity ($\text{m}\cdot\text{h}^{-1}$)
1	0.5	0.3	0.2	0.3
2	0.5	0.3	0.2	0.8
3	0.5	0.7	0.3	0.3
4	0.5	0.7	0.3	0.8
5	0.8	0.3	0.3	0.3
6	0.8	0.3	0.3	0.8
7	0.8	0.7	0.35	0.3
8	0.8	0.7	0.35	0.8
Base	0.7	0.5	0.25	0.5

6.4 Design Variable Justification

Evaporation fraction (α_{evap}) was varied because it directly controls the degree of solvent recovery in the recycle loop and hence the concentration of the mother liquor entering downstream crystallisation. This parameter has a first-order impact on crystallisation yield, purge losses, solvent consumption, PMI, and operating cost, making it a dominant flowsheet-level design decision [2][9][12].

Wash solvent mass fraction was included as it determines the extent to which impurity-rich mother liquor is displaced from the crystal cake during solid-liquid separation. In industrial crystallisation, washing is a primary lever for meeting product purity specifications but is also a major contributor to solvent use, waste generation, and energy demand. Varying the wash ratio therefore captures the trade-off between potential purity benefits and sustainability and cost penalties [2][9][21].

Solvent hold-up in the filter cake was varied to reflect the physical realities of filtration and washing, where a fraction of solvent remains entrained with the solids. Cake hold-up directly influences the amount of solvent forwarded to the recovery loop and thus affects PMI, energy duties, and equipment sizing. Including this variable ensures the separation model does not unrealistically assume ideal dewatering [20-28].

Filter superficial velocity (TEA_FI_v_sup_mph) was investigated because it affects the required filtration area and, consequently, capital cost in the techno-economic analysis. While it does not alter the underlying mass or energy balances, it represents a key engineering sizing assumption that can materially influence CapEx and annualised operating costs. Its inclusion allows the separation of flowsheet-level performance effects from equipment-level economic optimisation.

Moreover, a consistent set of performance indicators was evaluated to quantify the impact of design-variable changes on process performance, sustainability and

economics. These metrics include overall crystallisation yield, Process Mass Intensity (PMI), average E-factor, specific heating and cooling duties per unit mass of API, total capital expenditure (CapEx), and operating expenditure normalised per kilogram of API (OpEx). Tracking the same indicators for all cases enables direct comparison of material efficiency, energy intensity and economic performance, and provides a clear basis for identifying trade-offs and selecting the most favourable process configuration [2][6][17][21].

6.5 How to Vary Design Variables Chosen

The evaporation fraction (α_{evap}) is adjusted indirectly by changing flash-evaporator operating conditions. Higher values are achieved through increased heat input, reduced operating pressure (vacuum), or longer residence time, all of which promote methanol vaporisation. Lower values follow from gentler heating, higher pressure, or shorter residence times. The practical range is limited by solvent thermodynamics, utility availability, and API thermal stability [12][14][17].

The results for the cases obtained are as follows (*Table 6.5.1*):

Case	Operating Cost (£·kg ⁻¹ API)	Total Capital Cost (£)	Process Mass Intensity (PMI)	Average E-factor (-)	Overall Crystallisation Yield (-)	Specific Heating Duty (kJ·kg ⁻¹ API)	Specific Cooling Duty (kJ·kg ⁻¹ API)
1	4.055	3,603,617	4.805	2.762	0.638	2,935	-3,818
2	4.04	3,517,672	4.805	2.762	0.638	3,789	-3,818
3	4.499	3,818,555	5.274	3.525	0.581	3,789	-4,831
4	4.481	3,725,176	5.274	3.525	0.581	3,789	-4,831
5	3.087	3,786,329	3.598	1.105	0.852	4,238	-5,201
6	3.073	3,677,337	3.598	1.105	0.852	4,238	-5,201
7	3.284	4,166,158	3.745	1.361	0.819	5,899	-7,086
8	3.268	4,046,790	3.745	1.361	0.819	5,899	-7,086
Base	1.304	3,938,942	4.180	1.904	0.734	4,489	-287.8427

6.6 Effect of Design Variables

Effect of Evaporation Fraction (α_{evap})

Increasing the evaporation fraction from 0.50 to 0.80 strengthens solvent recovery and raises crystallisation yield in all tested configurations. However, this enhancement comes with a clear penalty: higher heating and cooling demands and greater operating costs per kilogram of API. Even at the best-performing high-evaporation condition (Case 6), operating cost (~£3.07/kg) remains more than twice that of the Base Case (£1.30/kg). Similarly, while Cases 5-6 achieve yields near 0.85, this improvement does not compensate for the substantially larger thermal and economic burdens compared with the Base Case yield of 0.73, achieved with much lower utility consumption [12,17].

The wash mass fraction is varied by altering the wash solvent flow rate, total wash volume, or number of wash stages applied to the cake. It can also be influenced by wash mode (spray, soak, reslurry) and contact time. Cake thickness affects the required wash ratio because thicker beds need more solvent to displace mother liquor effectively.

Cake solvent hold-up cannot be directly manipulated but depends on crystal properties and filtration operating conditions. Hold-up is reduced by producing larger, less compressible crystals, applying higher pressure or vacuum during filtration, extending drainage time, or using more permeable filter media. Gas blowing or mild heating can further enhance dewatering, while excessive pressure must be avoided to prevent cake compaction. Filter superficial velocity is a sizing decision governed by the ratio of slurry throughput to installed filter area. Higher velocities correspond to smaller filter areas or higher feed rates supported by sufficient pressure or vacuum. Lower velocities require larger filtration area or gentler pressure gradients. Practical limits are set by cake structure, slurry viscosity, and achievable pressure drop [24-28].

Effect of Wash Solvent Ratio and Cake Solvent Hold-up

Increasing wash ratio or solvent retention in the cake consistently worsens PMI, E-factor, and yield by enlarging the circulating solvent inventory and raising purge losses. Cases 3-4 and 7-8 illustrate this clearly: they show the highest PMI (up to 5.27) and E-factor (up to 3.53), paired with the lowest yields. In contrast, the Base Case attains a lower PMI (4.18) and E-factor (1.90) without the need for intensified washing, confirming that aggressive wash conditions introduce sustainability penalties without delivering proportional purity or recovery benefits [20-28].

Effect of Filter Superficial Velocity

Varying filtration velocity primarily alters equipment sizing and has negligible influence on material and energy flows. Although higher superficial velocity (e.g., Case 2 vs Case 1) slightly reduces CapEx, the magnitude of this reduction is minimal relative to total project cost,

and it does not improve process sustainability or reduce energy consumption. These equipment-level adjustments do not offer any advantage significant enough to surpass the strong overall performance of the Base Case.

6.7 Overall Comparison and Optimal Configuration

When viewed holistically, the Base Case (*Table 6.5.1*) emerges as the clear overall winner. It offers the lowest operating cost (£1.30/kg vs £3.07-4.50/kg across cases), avoids the high energy penalties associated with deeper solvent recovery, and does not require oversized equipment. Although the Base Case yield (0.73) is lower than the highest-yield scenarios, its drastically reduced utility demand and superior economic performance outweigh yield improvements achieved at the expense of energy intensity and cost [16][17].

Thus, while the parameter study clarifies how each design variable shapes mass efficiency, energy use, and economics, the unmodified Base Case provides the most industrially attractive solution, highlighting that aggressively intensifying solvent recycling and washing does not necessarily deliver an overall benefit.

This analysis can be replicated in industry on a qualitative basis since it mirrors the type of qualitative trade-off assessment routinely carried out in industrial process development, where yield improvements are balanced against utility demand, operability, and lifecycle costs.

6.8 Future Scope

The model assumes a fixed recycle composition and impurity load, which limits realism; allowing dynamic impurity build-up would improve accuracy. Solubility is represented by a single correlation, so using more rigorous or composition-dependent thermodynamics could refine predictions. Energy integration is simplified to one heat exchanger; exploring advanced heat-recovery strategies could reduce utility demand. The study is steady-state only; dynamic modelling would enable analysis of start-up, shutdown and control behaviour.

A broader optimisation framework, including purity constraints, Life Cycle Analysis metrics and improved global sensitivity analysis, would strengthen overall process design [3][4].

7 Conclusion

This work has developed and applied an integrated CAMPD-flowsheet-TEA framework for the continuous cooling crystallisation of paracetamol in methanol with solvent recycling. Using gPROMS coupled to SAFT- γ -Mie, a closed-loop process delivering 187.5 kg·h⁻¹ API ($\approx 1\%$ global demand) was modelled, sized and costed consistently from steady-state mass and energy balances. Capital costs were estimated via a Chilton-style cascade, while operating costs combined solvent, utility and labour charges into a single £·kg⁻¹ API metric aligned with PMI and E-factor. The system is projected to be profitable since its first year of operation, assuming the selling prices of paracetamol to be £5.23 kg⁻¹. At a 2% inflation rate, over a 20-year project basis with £5.5 million revenue in the first year, the lifetime revenue of the process is projected to be £134 million.

The results show that crystallisers dominate the capital expenditure cost, whereas solvent consumption and make-up govern OPEX. When various design options were compared, it became evident that increasing the evaporation fraction strengthens solvent recovery and yield but leads to disproportionately higher heating and cooling duties, ultimately leading to operating costs more than double those of the Base Case. Likewise, higher wash ratios and cake solvent hold-up systematically degrade PMI and E-factor by inflating circulating solvent inventory and purge losses. Variations in filtration velocity mainly influence filter area and have only marginal economic impact.

When all metrics are considered jointly, the unmodified Base Case emerges as the most attractive configuration, offering the lowest operating cost and competitive PMI/E-factor without excessive energy intensity or oversizing of equipment. The study therefore demonstrates that aggressive intensification of solvent recovery and washing does not guarantee superior techno-economic performance. The modelling and analysis protocol established here is directly transferable to other API-solvent systems and provides a structured basis for future extensions, including dynamic operation, enhanced heat integration and life-cycle-based sustainability assessment.

8 Acknowledgement

We would like to express our sincere gratitude to our project supervisors and the Department of Chemical Engineering for their guidance, expertise, and continuous support throughout this work. We are especially grateful to Dr Claire Adjiman and Dr Gaurav Seth, whose mentorship and encouragement were instrumental in the successful completion of this project. We also extend our heartfelt thanks to our families and friends for their unwavering encouragement during this journey.

References

- [1] Variankaval, N.; Cote, A. S.; Doherty, M. F. From Form to Function: Crystallization of Active Pharmaceutical Ingredients. *AIChE J.* 2008, 54 (7), 1682-1688. <https://doi.org/10.1002/aic.11555>
- [2] Jiménez-González, C.; Ponder, C. S.; Broxterman, Q. B.; Manley, J. B. Using the Right Green Yardstick: Why Process Mass Intensity Is Used in the Pharmaceutical Industry to Drive More Sustainable Processes. *Org. Process Res. Dev.* 2011, 15 (4), 912-917. <https://doi.org/10.1021/op200097d>
- [3] Burcham, C. L.; Florence, A. J.; Myerson, A. S. Continuous Crystallization in the Pharmaceutical Industry. *Annu. Rev. Chem. Biomol. Eng.* 2018, 9, 253-281. <https://doi.org/10.1146/annurev-chembioeng-060817-084055>
- [4] Mascia, S. et al. End-to-End Continuous Manufacturing of Pharmaceuticals: Integrated Synthesis, Purification, and Final Dosage Formation. *Angew. Chem. Int. Ed.* 2013, 52 (47), 12359-12363. <https://doi.org/10.1002/anie.201305429>
- [5] Diab, S., & Gerogiorgis, D. I. Technoeconomic Evaluation of Multiple Mixed Suspension-Mixed Product Removal (MSMPR) Crystallizer Configurations for Continuous Cyclosporine Crystallization. *Org. Process Res. Dev.* 2017, 21 (11), 1675-1683. <https://doi.org/10.1021/acs.oprd.7b00225>
- [6] Papaioannou, V.; Lafitte, T.; Avgousti, A.; Adjiman, C. S.; Jackson, G.; Galindo, A.; Müller, E. A. Group Contribution Methodology Based on the Statistical Associating Fluid Theory for Heteronuclear Molecules Formed from Mie Segments. *J. Chem. Phys.* 2014, 140, 054107. <https://doi.org/10.1063/1.4851455>
- [7] Granberg, R. A.; Rasmuson, Å. C. Solubility of Paracetamol in Pure Solvents. *J. Chem. Eng. Data* 1999, 44 (6), 1391-1395. <https://doi.org/10.1021/je990124v>
- [8] Mullin, J. W. *Crystallization*, 4th ed.; Butterworth-Heinemann: Oxford, 2001. <https://doi.org/10.1016/B978-0-7506-4833-2.X5000-0>
- [9] Capello, C.; Fischer, U.; Hungerbühler, K. What is a Green Solvent? A Comprehensive Framework for the Environmental Assessment of Solvents. *Green Chem.* 2007, 9 (9), 927-934. <https://doi.org/10.1039/B617536H>
- [10] Urwin, S. J.; Yerdelen, S.; Houson, I.; ter Horst, J. H. Impact of Impurities on Crystallization and Product Quality: A Case Study with Paracetamol. *Crystals* 2021, 11 (11), 1344. <https://doi.org/10.3390/cryst11111344>
- [11] Ledru, F.; Ruget, F.; Ventejou, B.; Desvignes, C. Study of the solid/liquid equilibrium of paracetamol in water, ethanol, and their binary mixtures. *Int. J. Pharm.* 2007, 339 (1-2), 32-38. <https://doi.org/10.1016/j.ijpharm.2007.02.016>
- [12] Green, D. W.; Southard, M. Z. *Perry's Chemical Engineers' Handbook*, 9th ed.; McGraw-Hill Education: New York, 2019.
- [13] Muhieddine, M. Model-Based Solvent Selection for Pharmaceutical Process Development. Ph.D. Thesis, Imperial College London, London, U.K., 2023.
- [14] Woods, D. R. *Rules of Thumb in Engineering Practice*; Wiley-VCH: Weinheim, 2007
- [15] *Methanol Market Report 2022–2023*; Independent Commodity Intelligence Services (ICIS): London, U.K., 2023.
- [16] Turton, R.; Bailie, R. C.; Whiting, W. B.; Shaeiwitz, J. A. *Analysis, Synthesis, and Design of Chemical Processes*, 4th ed.; Prentice Hall: Upper Saddle River, NJ, 2012
- [17] Peters, M. S.; Timmerhaus, K. D.; West, R. E. *Plant Design and Economics for Chemical Engineers*, 5th ed.; McGraw-Hill: New York, 2003
- [18] Towler, G.; Sinnott, R. *Chemical Engineering Design: Principles, Practice and Economics of Plant and Process Design*, 2nd ed.; Butterworth-Heinemann: Boston, 2012.
- [19] Department for Business, Energy & Industrial Strategy. *Quarterly Energy Prices: Industrial Electricity Prices in the UK*; U.K. Government: London, U.K., 2022
- [20] Gate Fees 2021/22 Final Report: Comparing the Costs of Alternative Waste Treatment Options; Waste and Resources Action Programme (WRAP): Banbury, U.K., 2022.
- [21] *The Sustainability Metrics*; Institution of Chemical Engineers (IChemE): Rugby, U.K., 2002
- [22] *Employee Earnings in the UK: 2023 Annual Survey of Hours and Earnings (ASHE)*; Office for National Statistics (ONS): Newport, U.K., 2023.
- [23] British National Formulary (BNF) (n.d.) *Paracetamol: medicinal forms*. Available at: <https://bnf.nice.org.uk/drugs/paracetamol/medicinal-forms/#oral-tablet> (Accessed: [10.12.2025]).
- [24] *Handbook of Industrial Drying*; Mujumdar, A. S., Ed.; Publisher: Place of Publication, Year.
- [25] *Heinkel Filtration Engineering Notes*; Heinkel Process Technology: Place of Publication, Year.
- [26] Richardson, J. F.; Harker, J. H. *Chemical Engineering, Volume 2*; Publisher: Place of Publication, Year.
- [27] Heinkel & Rosen. *Nutsche Filter Design Manual*; Heinkel Process Technology: Place of Publication, Year.
- [28] Rushton. *Nutsche Filter Handbook*; Publisher: Place of Publication, Year.

Benchmarking Automated, High-throughput Adams' Fusion Synthesis of IrO₂ OER Electrocatalysts

Chenyu Yan and Yatian Li

Department of Chemical Engineering, Imperial College London, UK

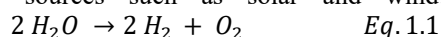
Abstract: This study benchmarks the impact of sodium nitrate (NaNO₃) loading and annealing conditions on iridium oxide (IrO₂) synthesised via an automated, high-throughput Adams' fusion workflow, enabling efficient future screening of lower Ir loading catalyst. Varying Ir precursor to NaNO₃ molar ratios demonstrated the feasibility of reducing NaNO₃ loading to improve the synthesis efficiency and sustainability, determined by the consistent rutile structure and high electrochemical performances. The optimised 50Na formulation was then used to investigate the annealing effects, where combined XRD, OER activity and ECSA identified 350 °C as the most favourable annealing temperature at this stage of evaluation. Overall, this work establishes a clear benchmark for key synthesis variables, validating the automated Adams' fusion as a robust and sustainable route to establish optimal conditions for future high-throughput studies on Ir-efficient OER catalysts.

Keywords: Adams' fusion, IrO₂ electrocatalyst, high-throughput workflow, automated synthesis, XRD, OER activity, ECSA, S-number, RDE, benchmark



1. Introduction

The transition towards low-carbon energy systems has increased the need for clean and economically viable energy systems¹. The UK government has legislated a set of regimes of achieving net-zero greenhouse gas emissions by 2050 to adhere to the goal set in the 2015 Paris Agreement². To comply with this commitment, hydrogen energy is highly recommended to mitigate the global reliance on fossil fuels, especially in the fields of various industries and vehicles³. Hydrogen emerges as a promising candidate due to its high gravimetric energy density and carbon-free utilisation⁴. Although several production pathways exist, most global hydrogen is still generated from fossil-fuel-based processes, such as via steam reformation. Green hydrogen produced via water electrolysis offers a sustainable alternative, shown below in Eq.1.1. This pathway relies only on water and generates only oxygen as a by-product, powered by renewable sources such as solar and wind.



There are three main water electrolysis technologies, Alkaline Water Electrolysis (AWE), Solid Oxide Water Electrolysis (SOWE), and Proton Exchange Membrane (PEM) electrolysis. Among these water electrolysis technologies, PEM electrolysis has emerged as a highly efficient and scalable option. Comparing the two most mature technologies, AWE and PEM, both typically exhibit Oxygen Evolution Reaction (OER) overpotentials of around 0.4 V; however, PEM electrolyzers operate at negligible Hydrogen Evolution Reaction (HER) overpotentials, showing a significant performance advantage in overall cell efficiency⁵. In addition, PEM electrolyzers are able to operate at high current densities with these comparatively overall lower overpotentials, delivering rapid dynamic response and maintain high stack efficiencies^{6,7}. To address the kinetically sluggish OER at the anode and its corresponding large overpotential, improving the electrochemical performance of the noble metal

electrocatalysts is essential to ensure system operational efficiency⁸.

Iridium oxide (IrO₂) is the state-of-the-art OER catalyst due to its exceptional activity and high corrosion resistance in acidic media⁹. However, the extreme scarcity and high cost of iridium impose significant constraints on large-scale commercial deployment¹⁰. The commonly used Ir precursor, iridium hexachloride hydrate (H₂IrCl₆·6H₂O), costs approximately £210 per gram, making it economically unsustainable for large-scale applications and highlighting the need for reducing Ir content.

To address the challenge of iridium scarcity, extensive research has focused on reducing Ir loading while retaining high OER performance. One of the most promising approaches is the development of Ir-based mixed-metal oxides (MMOs), where the IrO₂ lattice is partially substituted with more abundant acid-stable metals such as tantalum, titanium and tin¹¹. The insertion of these non-platinum group metals (non-PGMs) can substantially decrease Ir content while preserving or even enhancing the balance between activity and stability of the rutile structure¹². Additionally, Adams' Fusion¹³ can be reliably adopted as a robust synthesis route for the benchmarking and screening of the oxide electrocatalysts. In this synthesis, the chloride precursor of noble metal is substituted by sodium nitrate (NaNO₃) at elevated temperatures to produce rutile-phase metal oxide powders with high reproducibility and few defects¹⁴.

To further accelerate electrocatalyst discovery, automated high-throughput (HT) synthesis platforms have gained increasing attention, which are capable of screening multiple samples simultaneously with high reproducibility and increased cost-efficiency¹⁵. While automation significantly accelerates catalyst discovery, adapting

automation to the Adams' fusion method presents a practical challenge in NaNO_3 dispensing. The standard protocol uses a substantial excess of 5g NaNO_3 per 150 mg of iridium hexachloride hydrate ($\text{H}_2\text{IrCl}_6 \cdot 6\text{H}_2\text{O}$) to ensure complete reaction of the precious Ir precursor¹⁶. Such high NaNO_3 loading is incompatible with automated platforms, as slow dispensing rates (14 minutes per sample, exceeding 5 hours for a full batch of 24 samples) significantly increases synthesis time and reduces experimental throughput. Moreover, excessive NaNO_3 usage not only increases material consumption and operational cost, but also violates principles of green chemistry, particularly against waste minimisation and improved atom economy defined by the U.S. Environmental Protection Agency¹⁷. Optimising the Ir precursor to NaNO_3 molar ratio is therefore crucial to enable efficient automated synthesis and improve the overall sustainability, whilst maintaining the performance of electrocatalysts.

Following the optimisation of the reagent ratio, a systematic annealing study conducted in a muffle furnace is then carried out to benchmark the thermal post-treatment conditions for automated synthesis. This is because annealing temperature and duration directly affect the crystal structure of the IrO_2 catalyst, which influences the electrochemical performance significantly. Benchmarking annealing conditions allows the optimisation of IrO_2 catalytical performance alongside reduced energy requirements. With the successful establishment of this robust benchmark, multiple samples can be annealed simultaneously to improve the overall workflow efficiency and enable high-throughput electrocatalyst screening.

Therefore, this study aims to benchmark the adaptation of Adams' fusion synthesis to an automated high-throughput platform by (i) identifying the optimal Ir precursor to NaNO_3 molar ratio for the efficient and sustainable production of high-performance IrO_2 catalysts, and (ii) investigating the influence of annealing temperature and duration in a high-throughput muffle furnace to establish the optimal thermal conditions for generating highly active OER catalysts. Overall, this work provides a systematic framework for enhancing both synthesis efficiency and catalyst performance, and enabling rapid catalyst screening to support the development of next generation of a low-Ir catalyst used for PEM electrolyser technology.

2. Materials and Methodology

2.1. Preparation of IrO_2 Catalysts

Synthesis was performed on a Chemspeed FLEX ISYNTH platform, which is a 24 x 20 mL vials parallel reactor module, enabling efficient automated, high-throughput experimentation. Liquid reagents were delivered using the platform's

integrated volumetric unit, equipped with three independent 10 mL syringe channels and a separate channel for vacuum-assisted filtration. Solids were added gravimetrically via a GDU-P(fd) module featuring a 0.1 mg resolution balance. Reactions were carried out in sealed vials on a temperature-controlled agitation block (20 °C to 150 °C) fitted with individual reflux condensers to minimise solvent loss. The block provides uniform heating or cooling across all positions and programmable shaking to ensure consistent mixing. The entire workflow, which includes liquid and solid dispensing, homogenisation, thermal initiation, and solvent evaporation, was executed through fully automated scripts, thereby ensuring precise, reproducible synthesis conditions for all 24 parallel reactions.

A 30 mg mL^{-1} stock solution of hydrogen hexachloroiridate (IV) ($\text{H}_2\text{IrCl}_6 \cdot 6\text{H}_2\text{O}$) (110802-84-1, SIGMA-ALDRICH, USA) was dissolved in isopropyl alcohol (IPA) and loaded into the ISYNTH platform. 5 mL of this stock solution was dispensed using the automated needle head to each vial followed by an additional 5 mL of IPA to ensure all Ir precursor was delivered to each vial achieving a final Ir precursor concentration of 15 mg mL^{-1} . This was stirred at 300 rpm for 10 mins for even distribution. Sodium nitrate (NaNO_3) (7631-99-4, Sigma-Aldrich, India) was then gravimetrically dispensed using the automated GDU-P(fd) balance to achieve the desired Ir precursor to NaNO_3 molar ratios, followed by a further stirring at 300 rpm for 1 h. Next, the ISYNTH reactor block was heated to remove the IPA from all vials. This process resulted in dry precursor and salt mixtures.

The first part of the project aimed to determine the decreased optimal NaNO_3 loading required for a successful Adams' fusion synthesis¹³. The literature molar ratio 1:159 was selected as the reference point to assess both the influence of NaNO_3 quantity and the reliability of the ISYNTH platform. Two samples of 1:159 ratio was prepared using distinct dispensing methods: manual dispensing and automated gravimetric dispensing. In addition to this ratio, five further Ir precursor to NaNO_3 molar ratios, 1:1, 1:6, 1:8, 1:10, and 1:50 were prepared using the ISYNTH to evaluate the minimum NaNO_3 amount requirement for efficient dispensing time and successful synthesis. For the first study, dried precursor and salt mixtures were annealed in a tubular furnace at 450 °C for 1 h at a ramping rate of 10 °C min^{-1} . The selected annealing condition corresponds to the optimum identified in earlier optimisation work comparing treatments at 400 °C, 500 °C, and 600 °C for 1 h. This choice is also consistent with literature reports using annealing temperatures within the 350–600 °C range¹⁸. After annealing, the samples were cooled to room temperature, washed thoroughly with distilled water

to remove unreacted salts via centrifugation centrifuged (Eppendorf Centrifuge 5910 R_i) three times with deionised (DI) water, and dried on a hotplate. The resulting powders are labelled Auto_xNa or Man_xNa, where 'Auto' and 'Man' denote automated or manual NaNO₃ dispensing respectively, and x denotes the molar ratio of NaNO₃.

Using the optimum NaNO₃ amount, an annealing study was performed to benchmark a new muffle furnace for a more high-throughput approach. The muffle furnace allows multiple samples to be annealed simultaneously, improving the efficiency in the automated Adams' fusion workflow. Dried mixtures were annealed at 350 °C, 450 °C, and 550 °C for 1 and 2 hours, using the same heating rate of 10 °C min⁻¹. After thermal treatment, the samples were washed and dried as described above, yielding the final IrO₂ powders for characterisation and electrochemical testing. These samples follow the naming convention of temperature_time, for example 350 °C_1h.

2.2. Electrochemical Analysis

2.2.1. OER studies using liquid cell

All oxygen evolution reaction (OER) measurements were performed using a potentiostat (BioLogic VSP-3e) with a rotating disk electrode (RDE; Pine Research WaveVortex). A three-electrode configuration was employed to measure the activity and stability of the catalysts. The operating working electrode was generated by dropping a 0.2 mg mL⁻¹ catalyst ink onto a gold electrode insert (5 mm diameter and polished surface), the counter electrode was platinum mesh, and the reference electrode was a reversible hydrogen electrode. The electrolyte used was 0.1 M HClO₄ prepared from perchloric acid (70%, 017-006-00-4, Honeywell, Germany) and deionised water.

2.2.2. Working electrode preparation

The catalyst ink was prepared by dispersing catalyst powder in a mixed solvent of IPA and DI water (80:20 v/v) to obtain a final concentration of 0.2 mg mL⁻¹. This suspension was then sonicated for 10 minutes to ensure a uniform dispersion, after which 15 µL of Nafion (5 %, 31175-20-9, FUJIFILM, Germany) was added as a binder and sonicated for a further 10 minutes. For electrode preparation, 10 µL of 0.2 mg mL⁻¹ ink was drop-cast onto a gold electrode insert and dried at room temperature under a glass beaker for 2 h to promote even particle distribution.

2.2.3. Electrochemical performance

Electrochemical performance consisted of electrochemical impedance spectroscopy (EIS), cyclic voltammetry (CV), and OER activity measurements using the rotating disk electrode

(RDE). EIS was performed first to determine the uncompensated resistance (R_u), following a paper showing that accurate iR-correction improves the reliability of OER overpotential measurements¹⁹. The measured R_u was then applied as an 85% online ohmic correction during all CV scans. Next, cyclic voltammograms (CVs) were performed within a potential window free of charge-transfer reactions (0.8-1.2 V_{RHE})²⁰ at various scan rates (20, 50, 100, 120, 160 mV s⁻¹). These were used to evaluate the electrochemical double-layer capacitance C_{dl} , which serves as a prominent quantitative indicator of the accessible electrochemically active surface area (ECSA). For electrocatalysts, C_{dl} is typically extracted from the linear dependence of the capacitive current, defined as the difference (ΔI) between anodic and cathodic current in the non-Faradaic region on the scan rate v , where the slope of the ΔI versus v plot yields C_{dl} . Once C_{dl} is obtained, the ECSA can be quantified using $ECSA = C_{dl} / C_s$, where C_s is the specific capacitance of a flat reference surface. Literature values for C_s vary depending on the material, for example, IrO_x nanoparticles exhibit a reported C_s of 350 µF cm⁻²²¹⁻²³. Using the appropriate C_s allows accurate normalisation of C_{dl} and thus reliable estimation of the electrocatalyst's ECSA. Finally, OER activity was measured by CV, with scans from 1.20 to 1.65 V_{RHE} at 20 mV s⁻¹. The extracted data was used to calculate the current density (J_{geo} , mA cm⁻²), which acted as the activity comparison, as normalising the measured OER current (I_{OER} , mA) to the surface area of the RDE tip (A_{geo} , cm²), provides a consistent basis for evaluating the catalytic response of different samples under the same applied potential. All these measurements were performed with the working electrode rotating at 1600 rpm to ensure negligible mass-transport limitation.

The stability of the catalysts was assessed by extended CV cycling performed between 1.20 and 1.60 V_{RHE} at 50 mV s⁻¹ at 1600 rpm for 3000 cycles under room temperature, with 85% ohmic correction applied. After cycling, 6 mL of the electrolyte was collected to quantify the concentration of dissolved Ir using inductively coupled plasma mass spectrometry (ICP-MS). Geiger et al. introduced the stability number (S-number) as an intrinsic parameter for catalysts, defined as the ratio between the amount of oxygen generated and the amount of metal dissolved, which is a valuable tool for understanding and assessing the effectiveness of OER catalysts²⁴. The moles of evolved oxygen (n_{O_2}) were determined by integrating the anodic current recorded during the potential or current hold, using Eq. 2.2.3.1, where $z=4$ is the number of electrons transferred in the OER, F is the Faraday constant, and $i(t)$ is the measured OER current. The amount

of dissolved Ir, n_{Ir} , was quantified using an Agilent 8900 triple-quadrupole ICP-MS operated in MS/MS mode to eliminate spectral interferences and accurately determine ppb level Ir concentrations as shown in Eq. 2.2.3.2, where $[Ir]$ is the measured Ir concentration, V is the electrolyte volume, and M_{Ir} is the molar mass of Ir. The S-number was then obtained as Eq. 2.2.3.3 shows below, providing a quantitative measure of catalyst stability during oxygen evolution.

$$n_{O_2} = \frac{1}{zF} \int i(t) dt \quad Eq. 2.2.3.1$$

$$n_{Ir} = \frac{[Ir] \times V}{M_{Ir}} \quad Eq. 2.2.3.2$$

$$S - number = \frac{n_{O_2}}{n_{Ir}} \quad Eq. 2.2.3.3$$

2.3. Physical Characteristics

X-ray diffraction (XRD) patterns were performed by an X-ray diffractometer (Malvern PANalytical Aeris) equipped with a Cu $K\alpha$ radiation source ($K\alpha_1 = 1.5406 \text{ \AA}$, $K\alpha_2 = 1.5444 \text{ \AA}$, intensity ratio = 0.5) operating at 40 kV and 15 mA. A fixed divergence slit of 0.38 mm was used. Data were collected in continuous scan mode over a 2θ range of 20° – 60° , with a step size of 0.0109° and a counting time of approximately 53 s per step. The obtained diffraction patterns were used to analyse the crystal structure of the catalysts. X-ray photoelectron spectroscopy (XPS) was employed to analyse the near-surface elemental composition of the IrO_2 synthesised samples. Measurements were performed using a Thermo Fisher K-Alpha+ spectrometer equipped with a monochromatic Al $K\alpha$ X-ray source (1486 eV).

3. Results and Discussion

3.1. Study of Ir Precursor to $NaNO_3$ Mole Ratio on IrO_2

This study benchmarks the impact of $NaNO_3$ loading on IrO_2 synthesised via automated Adams' fusion. The optimal $NaNO_3$ loading is selected by considering the structural and electrochemical properties of the resulting IrO_2 , as well as the yield of the target product. This optimised loading is then used for the subsequent annealing study.

3.1.1. XRD Analysis

To understand how the synthesis parameters influence the final products, the structural properties of the IrO_2 samples were first examined. XRD analysis was used to identify the crystalline phases present and to evaluate how the precursor ratio affects material formation.

Fig.1 presents the XRD patterns of IrO_2 catalysts synthesised via the Adams' fusion method using

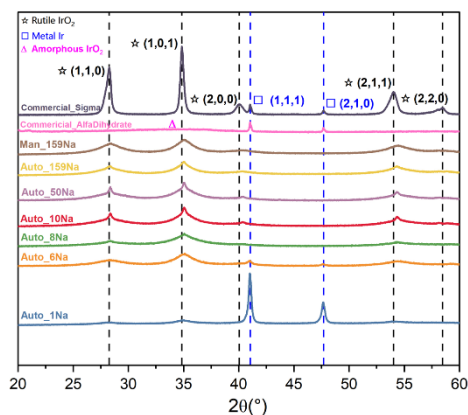


Fig. 1. XRD patterns of commercial IrO_2 and synthesised IrO_2 samples prepared under various $NaNO_3$ loading. Vertical dashed lines indicate the characteristic reflections of rutile IrO_2 , metallic Ir, and amorphous IrO_2 .

different $NaNO_3$ loadings, corresponding to Ir precursor to $NaNO_3$ molar ratios ranging from 1:1 to 1:159 and all samples were annealed at 450°C for 1h. The 2θ positions of the characteristic rutile structures are consistent with those of the commercial reference (Commercial_Sigma), where no peak shift is observed. This demonstrates there is no lattice distortion or metal incorporation across the 8Na to 159Na series. The structural consistency shown by automatically and manually synthesised samples (Auto_159Na and Man_159Na) successfully proves the feasibility of adapting Adams' fusion method to the automated high-throughput system.

For samples Auto_8Na to Man_159Na, the diffraction peaks are well indexed to the rutile crystallographic structure of IrO_2 , characterised by the enhanced intensity of preferential orientation (1,1,0) and (1,0,1) rutile IrO_2 crystal¹⁴. Rutile IrO_2 is the thermodynamically stable phase under acidic PEM electrolyser conditions that ensures the stability and durability of the electrocatalyst²⁵. Additional minor peaks in the orientations (2,0,0), (2,1,1) and (2,2,0) are also observed in the XRD analysis, further confirming the formation of rutile-structured IrO_2 . The absence of metal Ir peaks proves the complete reaction of the Ir precursor under these synthesis conditions.

The sample prepared with insufficient $NaNO_3$ exhibits weak and poorly defined IrO_2 diffraction peaks. Sample Auto_1Na was synthesised with $NaNO_3$ loading lower than stoichiometric ratio. Strong and sharp metallic Ir peaks at orientations (1,1,1) and (2,1,0) are clearly observed, which indicates incomplete conversion of the Ir precursor. Moreover, Auto_6Na was theoretically prepared at the exact stoichiometric ratio, but dispensing inaccuracies in the automated system resulted in an effective $NaNO_3$ loading that was lower than expected. As a result, although the sample displays broad rutile IrO_2 peaks at the (1,1,0), (1,0,1) and

(2,1,1) planes, additional minor diffractions feature corresponding to metallic Ir at (1,1,0) and (2,1,0) planes are also observed. The presence of these metal Ir signatures again confirms that a minimum stoichiometrically matched amount of NaNO_3 loading is required to fully drive the Adams' fusion reaction.

In contrast, samples prepared with NaNO_3 loadings from 1:8 to 1:159 (Auto_8Na to Auto_159Na) display similar peak widths and intensities, indicating similar crystallinity and confirming that reducing NaNO_3 to the minimum loading limit does not alter the result of crystal formation of the electrocatalyst. Therefore, the NaNO_3 loading can be reduced without compromising crystal structure and catalyst purity. This structural consistency across the NaNO_3 series is highly relevant for electrochemical performance, as electrochemically active surface area, reaction accessibility, and ultimately OER activity and stability can be greatly affected by the nanoscale crystallinity. Therefore, further assessments of the electrochemical performance and catalyst yield are essential to identify an optimised precursor to NaNO_3 ratio.

3.1.2. Electrochemical Performance

The OER activity of the IrO_2 catalysts synthesised with different Ir precursor to NaNO_3 ratios was assessed using rotating disk electrode (RDE) voltammetry in 0.1 M HClO_4 electrolyte, extracted data were based on the 20th CV cycle from total 26 cycles, which represents the stabilised electrochemical behaviour after initial activation. For comparison across samples, the current density J_{geo} at a fixed potential of 1.65 V is a representative point in the active region, as Fig.2(a) shown. The activity values present correspond to the average of two independent repeats, and the associated error bars reflect the variability between these measurements. The two commercial IrO_x samples were used as performance benchmarks to compare with the synthesised catalysts. The NaNO_3 loadings of the Auto_1Na and Auto_6Na samples were insufficient to complete the reaction, leading to the presence of metallic Ir or partially reacted species rather than pure IrO_2 as indicated by XRD results. Although metallic Ir can contribute to OER activity, its performance is markedly lower than IrO_2 and does not reflect the behaviour of the targeted catalyst. For the samples 8Na, 10Na, 50Na, and 159Na which have sufficient NaNO_3 loadings, although small variations in the averaged J_{geo} are observed, the overlap in the error bars shows that these differences lie within experimental uncertainty. Their comparable structures help explaining the similar electrochemical responses observed across the series. Overall, within the tested range, NaNO_3 loadings above the stoichiometric requirement did not affect the OER activity of the IrO_2 .

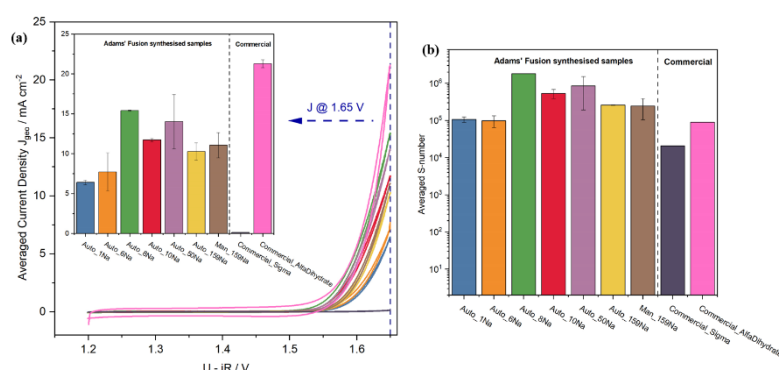


Fig. 2. Electrochemical performance of IrO_2 catalysts synthesised with different NaNO_3 loadings. (a) Averaged CV curves with the corresponding J_{geo} at 1.65 V shown in the inset. (b) Averaged S-number values reflecting catalyst stability. Synthesised samples are shown on the left of each panel, with commercial references displayed on the right.

Effective Ir-based OER catalysis requires not only high activity, but also by high stability under strongly oxidative and acidic operating conditions. The stability number (S-number) is a quantitative indicator for catalyst's stability, defined as the ratio between the amount of oxygen generated and the amount of iridium dissolved in the electrolyte. The higher the number, the more stable the electrocatalyst. As shown in Fig.2(b), Auto_1Na and Auto_6Na which contain metallic Ir show lower stability than others, as metallic Ir is known to be far less stable than IrO_2 under OER conditions²⁶. All catalysts synthesised with sufficient NaNO_3 ($\geq 8\text{Na}$) exhibit high S-numbers on the order of 10^5 - 10^6 , demonstrating excellent intrinsic stability even compared with the commercial catalysts. The absence of a systematic trend with increasing NaNO_3 loading, together with overlapping uncertainties, indicates that once the minimum NaNO_3 loading needed for full precursor reaction is supplied, further increases in NaNO_3 do not influence the dissolution behaviour of the resulting IrO_2 . This stability invariance is also consistent with the XRD results, which shows all fully reacted samples possess nearly identical rutile IrO_2 diffraction patterns, suggesting similar crystallinity and catalyst morphology. The comparison with commercial benchmarks further shows the advantages of the Adams' fusion route, as both commercial references exhibit lower stability than the synthesised samples under identical testing conditions. In addition, the manually prepared 159Na sample displays a stability comparable to that of the automatically dispensed 159Na sample, demonstrating that the Chemspeed ISYNTH platform does not compromise catalyst durability and is capable of reliably reproducing the performance obtained through manual synthesis.

In summary, the OER activity values fall within experimental uncertainty, and the S-number analysis shows comparable resistance to iridium dissolution for all adequately reacted samples. Together, these

results indicate that once sufficient NaNO_3 is present to complete the Adams' fusion reaction, increasing the NaNO_3 amount does not significantly influence the electrochemical behaviour or stability of the resulting IrO_2 . This finding allows the NaNO_3 loading to be reduced relative to the large excess typically recommended in the literature, without sacrificing catalyst quality, offering a practical route to shorten dispensing time from 14 mins (Auto_159Na) to 4 min (Auto_50Na), lower reagent use, and improve the efficiency of the automated synthesis workflow.

3.1.3. Yield Analysis

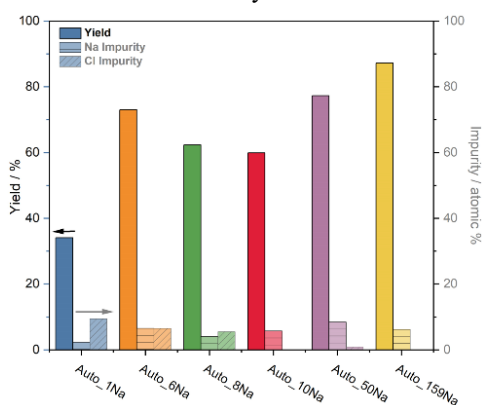


Fig. 3. Yield of the automatically synthesised IrO_2 samples alongside Na and Cl impurity amounts measured by XPS. Yield values are plotted on the left axis, while impurity contents are shown on the right axis.

However, synthesis yield and catalyst purity are also critical factors to be taken into account. Fig.3 summarises the isolated yields of the electrocatalysts synthesised across different Ir precursor to NaNO_3 molar ratios and Na and Cl impurities, characterised by atomic percent. The yields are corrected by excluding the mass of Na and Cl impurities, ensuring the reported values reflect only the mass of IrO_2 produced. After thorough catalyst washing, XPS analysis confirms that the residual Na and Cl contents from NaNO_3 and NaCl remained consistently below 10 wt% for all samples. The low impurity levels are attributed to the high solubility of NaCl and NaNO_3 in DI water, which allows efficient removal during the washing process. Notably, there is no clear trend in Na contamination observed across the different molar ratios, indicating that variations in NaNO_3 loading do not significantly influence impurity presence under the applied washing protocol. The yield of samples synthesised from precursor amount under stoichiometric ratio (Auto_1Na) is observed to be significantly lower than others. This again proved the insufficient reaction of the Ir precursor. The sample synthesised with the highest NaNO_3 loading (Auto_159Na) exhibited the highest yield, which is consistent with the result of complete reaction in Adams' fusion synthesis. In specific, Auto_50Na showed approximately 30 % higher yield than Auto_8Na and

Auto_10Na. The reduced yields observed for Auto_8Na and Auto_10Na suggest possible incomplete precursor conversion at these stoichiometries. This could be due to inability for the precursor to access sufficient NaNO_3 within the ISYNTH set up, therefore, higher NaNO_3 contents are more favourable for higher yields.

From a practical perspective, although Auto_6Na shows a relatively high measured yield, its reliability remains uncertain due to dispensing inaccuracies of the automated high-throughput system. The effective NaNO_3 loading is lower than intended, leading to incomplete precursor conversion. As a result, the yield highly possibly consists of a mixture of IrO_2 produced and unreacted or partially reacted Ir species. This interpretation is consistent with the previous electrochemical analysis, in which Auto_6Na shows lower activity compared to other fully reacted samples. Considering the high cost of Ir precursor and the scarcity of Ir metal, ensuring complete reaction is critical to achieve both economic and material sustainability. Auto_50Na is therefore revealed as a more optimal choice, stemming from its balance of a high yield whilst only required a third of NaNO_3 compared to the literature loading (Auto159Na).

Overall, balancing out structural characteristics, electrochemical performance, reagent efficiency and yield, the molar ratio of 1:50 (Auto_50Na) was therefore selected as the optimal ratio for subsequent annealing studies.

3.2. Study of Different Annealing Conditions on IrO_2 Phase

This study provides a robust benchmark on the impact of annealing conditions on the crystallinity, OER performance and yield of the IrO_2 synthesised with the 1:50 Ir precursor to NaNO_3 molar ratio. Through the analysis of a combination of XRD, activity and ECSA, a clear structure-property relationship is revealed, which enables reproducible thermal treatment conditions for future high-throughput catalyst screening.

3.2.1. XRD Analysis

Fig.4 summarises the XRD results of the IrO_2 catalysts synthesised all with precursor ratio 1:50, annealed at 350-550 °C for 1 h and 2 h. Across all samples, the diffraction patterns display no metallic Ir peaks, and no shifts in the 2θ positions of the IrO_2 peaks compared to the commercial samples.

At the lowest annealing temperature investigated, both 350 °C_1h and 350 °C_2h exhibit a single broad peak at the orientation (1,0,1), which reveals the characteristic of an amorphous and disordered IrO_2 structure. This broad peak greatly aligns with the trend shown in the commercial reference

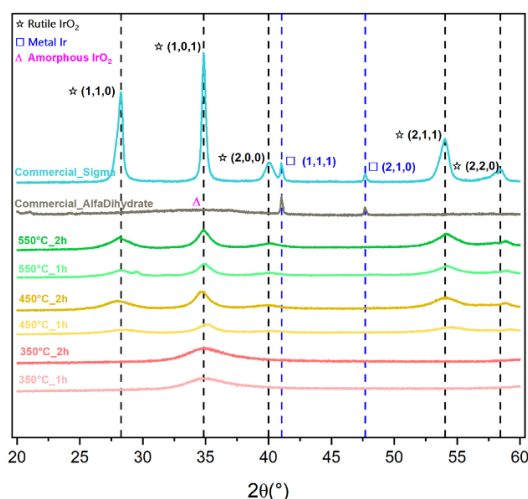


Fig. 4. XRD patterns of commercial IrO_2 samples and synthesised IrO_2 annealed at 350 °C, 450 °C, and 550 °C for 1 h and 2 h. Vertical dashed lines correspond to the characteristic reflections of rutile tetragonal IrO_2 . Markers indicate the presence of metallic Ir and amorphous IrO_2 .

Commercial AlfaDihydrate. Although the annealing temperature at 350 °C is observed to be insufficient to drive highly ordered crystallisation within 1-2 h timeframe, high catalytical activity could be expected, stemming from the high catalytical surface area provided by the amorphous structure.

Increasing the annealing temperature to 450 °C promoted the onset of rutile structure formation. Sample 450 °C_1h shows weak but distinguishable peaks corresponding to the (1,1,0), (1,0,1) and (2,1,1) planes, potentially indicating the presence of a phase in between amorphous and rutile structure. Extending the annealing time to 450 °C_2h results in the production of a clearer rutile structure due to its sharper peaks, confirming that both annealing temperature and time are critical determining factors of catalyst crystallinity.

Annealing temperature at 550 °C produced IrO_2 catalyst with increased rutile-structured crystallinity. For both 550 °C_1h and 550 °C_2h, characteristic diffractions are indexed to the (1,1,0), (1,0,1), (2,0,0), (2,1,1) and (2,2,0) planes. The similarity in diffraction between 550 °C_1h and 550 °C_2h indicates that, at higher temperatures, annealing time imposes less significant effects on crystallite formation.

Overall, the above XRD analysis demonstrates a temperature-driven progression from amorphous IrO_2 annealed at 350°C to rutile IrO_2 annealed at 450-550 °C. The crystallinity of the electrocatalyst increases with the elevated annealing temperature, which could potentially reduce electrochemical surface area and activity but improve structural stability and catalytical durability. These observations provide an essential structural basis for

correlating annealing conditions with catalytic performance in subsequent electrochemical analysis.

3.2.2. Yield Analysis

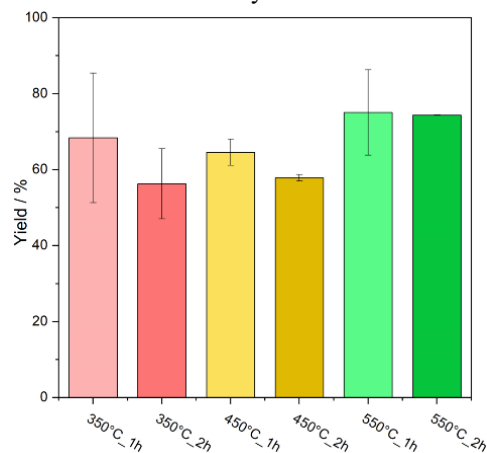


Fig. 5. Effect of the annealing temperature and time on IrO_2 synthesised yield. Samples treated at 350 °C, 450 °C, and 550 °C for 1 h and 2 h with measured uncertainties shown as error bars.

Fig.5 summarises the yields of the Ir precursor to NaNO_3 molar ratio of 1:50 samples annealed at 350-550 °C for 1 h and 2 h. Overall, there are no substantial differences observed in yield across these annealing conditions. This indicates that once the precursor was fully reacted, subsequent thermal post-treatment does not have a large impact on the total mass of IrO_2 obtained. Although a consistent trend of slightly lower yield for the 2 h samples is observed at all temperatures, the magnitude of this decrease is within the experimental uncertainty of the measurement. Therefore, the trend cannot be taken as definitive evidence of the impact of annealing durations.

Overall, while 1h annealing appears adequate from a material and energy efficiency perspective, the choice between 1h and 2h should ultimately be guided by further electrochemical performance. Further activity, ECSA and stability analysis are required to determine intrinsic impact of the annealing conditions on catalytic behaviour.

3.2.3. Electrochemical Performance

The OER activity of the 50Na series annealed at different temperatures and durations is shown in Fig.6 (a), displays a clear dependence on the thermal treatment. Samples annealed at 350 °C at both 1 h and 2 h give the highest current densities

compared to all other synthesised samples and the Commercial AlfaDihydrate reference. Raising the annealing temperature to 450 °C leads to a large drop in activity, and at 550 °C, the OER response is almost absent under the same testing conditions. These results aligns with the expectation drawn from the variation in structure with heating. Higher temperatures lead to a more crystalline and ordered catalyst²⁷. From the XRD results, 350 °C samples

show broad, low-intensity peaks, indicating a more amorphous characteristic, while 450 °C and 550 °C samples show sharper peaks which matching with rutile IrO_2 . This shift from a more disordered material at 350 °C to well crystallised rutile at higher temperatures provides a reasonable explanation for the loss of activity with increased annealing temperature and duration.

J_{geo} provides a convenient initial comparison of OER activity, however, it does not reveal how much of the catalyst surface is involved in the reaction. This limitation is relevant for IrO_2 nanoparticles, where only a fraction of catalyst active sites is accessible to the electrolyte, with electrochemical reactions occurring exclusively here. To gauge the accessible fraction, the electrochemical surface area (ECSA) was determined. ECSA estimates the surface area that is electrochemically active, the interface over which charge-transfer reactions, thus OER, can occur. Typically, larger ECSA values imply a greater population of active sites, higher active surface area and a higher capacity for reaction rate²⁸. In this study, ECSA is reported as supporting information rather than used for activity normalisation as it provides valuable insights on the surface characteristics of the synthesised catalysts and helps assess whether variations in annealing conditions influence their electrochemical accessibility.

The ECSA trends (Fig.6 (b)) mirror the temperature-dependent behaviour seen in the activity data. The 350 °C samples display the highest ECSA, indicating the greatest number of electrochemically accessible surface sites. ECSA falls sharply at 450 °C and becomes very low at 550 °C, consistent with the XRD evidence for progressive crystallite growth and reduced disorder at higher annealing temperatures. The commercial references also provide a useful contrast. Commercial_Sigma shows negligible ECSA, consistent with its poor activity whereas the Commercial_AlfaDihydrate exhibits the highest ECSA leading to its high activity and amorphous structure. One exception is that 350 °C_2 h has a slightly higher ECSA but a lower OER current density than 350 °C_1 h. This discrepancy suggests that not all electrochemically accessible surface area contributes equally to catalytic performance²⁹.

For the 50Na series, 350 °C is identified as the optimal annealing temperature thus far, providing the highest OER activity and the greatest electroactive surface area. Annealing at higher temperature (450 to 550 °C) reduce surface area and activity. The 350 °C_1 h and 350 °C_2 h samples perform equivalently within experimental error. However, no systematic effect of annealing duration can be concluded based on the limited dataset. Nevertheless, stability measurements must be performed to determine the optimal annealing

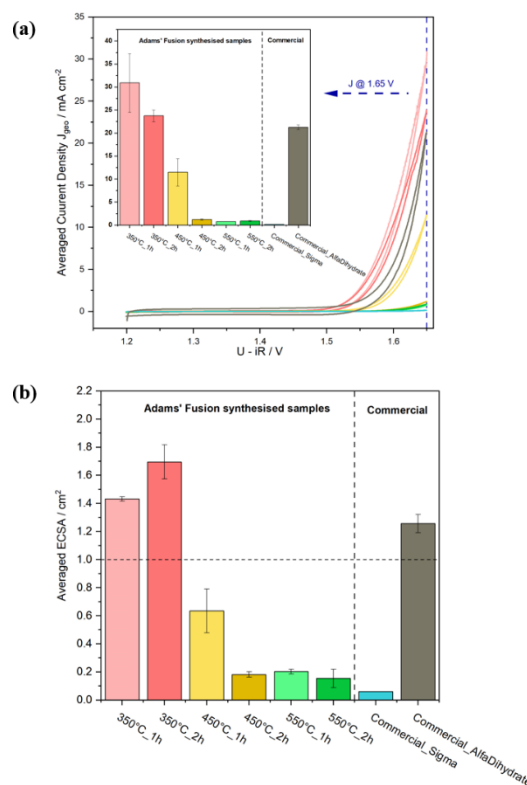


Fig. 6. Electrochemical performance of IrO_2 samples synthesised at different annealing temperatures and durations. (a) CV graph for averaged OER current density (J_{geo}) as a function of voltage, with inset showing averaged J_{geo} at 1.65 V. (b) Averaged electrochemically active surface area (ECSA) for the same samples, compared to the commercial references

conditions. Due to time constraints, this was not achieved for this project.

4. Conclusion

This study provides a systematic investigation into how NaNO_3 loading and post-synthesis thermal treatment affect the structure and electrochemical performance of IrO_2 produced via an automated, high-throughput Adams' fusion process. By evaluating a broad range of Ir precursor to NaNO_3 ratios, the results clearly show that once the minimum amount of NaNO_3 required for full precursor conversion is supplied, any further increase does not significantly alter the resulting rutile IrO_2 structure, its active surface area, and its OER electrochemical performance. XRD results confirmed consistent rutile IrO_2 formation across all sufficiently reacted samples, and electrochemical measurements (OER activity and S-number) were consistent within experimental uncertainty. This finding enables a substantial reduction in NaNO_3 usage compared with traditional literature practice, improving both the efficiency and sustainability of high-throughput synthesis and reducing the bottleneck associated with slow automated dispensing. Using the optimised 50Na formulation, the annealing study further revealed that temperature is a critical parameter governing catalyst

performance. The XRD, activity and ECSA results identify 350 °C as the temperature which gives the most active catalyst. At higher annealing temperatures (450-550 °C), crystallinity increases while surface accessibility decreases, and OER performance declines rapidly. Long-term dissolution stability was not assessed, but the structural and electrochemical trends indicate that higher temperature or longer duration anneals reduce the population of active surface features. The automated Adams' fusion workflow reproducibly produces high quality IrO₂ while reducing NaNO₃ use and enabling efficient thermal optimisation. These findings support more sustainable and scalable synthesis and provide a basis for future work in mixed-metal oxides synthesis to lower Ir content.

5. Outlook

The findings of this study point to several clear next steps for improving the automated Adams' fusion process and adapting it for future producing high-performing iridium-based catalysts. Reducing the NaNO₃ loading without loss of performance offers a practical route to improving the sustainability and throughput of the synthesis, particularly by shortening dispensing times on the Chemspeed FLEX ISYNTH platform. Completing the stability assessment of furnace annealed samples through ICP-MS dissolution measurements will be an important next step to confirm whether the improved activity and surface accessibility observed at 350 °C are accompanied by favourable long-term stability. In addition, advanced characterisation like transmission electron microscopy (TEM) would provide deeper insights. TEM could reveal particle size and crystallite structure that are not visible in XRD data. This would be especially useful for distinguishing between samples with similar electrochemical performance but potentially different nanoscale characteristics³⁰. An important opportunity for expansion lies within the DIGIBAT laboratory, where another Chemspeed platform is equipped with an automated annealing module capable of executing thermal treatments through programmable, code-based workflows. Additionally, integrating this system with the ISYNTH workflow would automate both synthesis and annealing, enabling fully high-throughput operation and faster optimisation and screening. Coupling the setup with rapid electrochemical tests and data-driven optimisation would further speed candidate evaluation, for example, a multi-working electrode used for OER testing. Finally, applying the optimised workflow to Ir-efficient catalyst designs such as mixed-metal oxides or supported catalysts represents an important step toward developing economically and materially sustainable OER catalysts for PEM water electrolysis.

Acknowledgements

The authors would like to thank Prof. Camille Petit and her PhD student Morven Holland for their invaluable guidance, insightful technical discussions, and generous support throughout this work. We also thank them for providing essential access to their laboratory facilities.

References

1. Gür, T. M. Review of electrical energy storage technologies, materials and systems: challenges and prospects for large-scale grid storage. *Energy Environ. Sci.* **11**, 2696–2767 (2018).
2. Horowitz, C. A. Paris Agreement. *Int. leg. mater.* **55**, 740–755 (2016).
3. Shahzad, S., Faheem, M., Muqet, H. A. & Waseem, M. Charting the UK's path to net zero emissions by 2050: Challenges, strategies, and future directions. *IET Smart Grid* **7**, 716–736 (2024).
4. Singla, M. K., Nijhawan, P. & Oberoi, A. S. Hydrogen fuel and fuel cell technology for cleaner future: a review. *Environ Sci Pollut Res* **28**, 15607–15626 (2021).
5. Bernt, M., Siebel, A. & Gasteiger, H. A. Analysis of Voltage Losses in PEM Water Electrolyzers with Low Platinum Group Metal Loadings. *J. Electrochem. Soc.* **165**, F305–F314 (2018).
6. Shiva Kumar, S. & Himabindu, V. Hydrogen production by PEM water electrolysis – A review. *Materials Science for Energy Technologies* **2**, 442–454 (2019).
7. Wang, T., Cao, X. & Jiao, L. PEM water electrolysis for hydrogen production: fundamentals, advances, and prospects. *Carb Neutrality* **1**, 21 (2022).
8. Fabbri, E., Haberer, A., Waltar, K., Kötz, R. & Schmidt, T. J. Developments and perspectives of oxide-based catalysts for the oxygen evolution reaction. *Catal. Sci. Technol.* **4**, 3800–3821 (2014).
9. Carmo, M., Fritz, D. L., Mergel, J. & Stolten, D. A comprehensive review on PEM water electrolysis. *International Journal of Hydrogen Energy* **38**, 4901–4934 (2013).
10. Minke, C., Suermann, M., Bensmann, B. & Hanke-Rauschenbach, R. Is iridium demand a potential bottleneck in the realization of large-scale PEM water electrolysis? *International Journal of Hydrogen Energy* **46**, 23581–23590 (2021).
11. Wang, H. *et al.* Recent Progress in Balancing the Activity, Durability, and Low Ir Content for Ir-Based Oxygen Evolution Reaction Electrocatalysts in Acidic Media. *Small* **21**, 2410407 (2025).
12. Over, H. *et al.* Hydrogenation of Mixed Ir-Ti Oxide, a Powerful Concept to Promote the

- Oxygen Evolution Reaction in Acidic Water Electrolysis. Preprint at <https://doi.org/10.26434/chemrxiv-2025-0fp67> (2025).
13. Shriner, R. L. & Adams, R. THE PREPARATION OF PALLADOUS OXIDE AND ITS USE AS A CATALYST IN THE REDUCTION OF ORGANIC COMPOUNDS.¹ VI. *J. Am. Chem. Soc.* **46**, 1683–1693 (1924).
 14. Siracusano, S., Van Dijk, N., Payne-Johnson, E., Baglio, V. & Aricò, A. S. Nanosized IrO_x and IrRuO_x electrocatalysts for the O₂ evolution reaction in PEM water electrolyzers. *Applied Catalysis B: Environmental* **164**, 488–495 (2015).
 15. Dai, T. *et al.* Autonomous mobile robots for exploratory synthetic chemistry. *Nature* **635**, 890–897 (2024).
 16. Felix, C., Maiyalagan, T., Pasupathi, S., Bladergroen, B. & Linkov, V. Synthesis and Optimisation of IrO₂ Electrocatalysts by Adams Fusion Method for Solid Polymer Electrolyte Electrolysers. *MNS* **4**, 186–191 (2012).
 17. Sheldon, R. A. Fundamentals of green chemistry: efficiency in reaction design. *Chem. Soc. Rev.* **41**, 1437–1451 (2012).
 18. Abbott, D. F. *et al.* Iridium Oxide for the Oxygen Evolution Reaction: Correlation between Particle Size, Morphology, and the Surface Hydroxo Layer from Operando XAS. *Chem. Mater.* **28**, 6591–6604 (2016).
 19. Wu, C. *et al.* Recommended electrochemical measurement protocol for oxygen evolution reaction. *DeCarbon* **8**, 100108 (2025).
 20. Wei, C. *et al.* Approaches for measuring the surface areas of metal oxide electrocatalysts for determining their intrinsic electrocatalytic activity. *Chem. Soc. Rev.* **48**, 2518–2534 (2019).
 21. Sun, W. *et al.* A promising engineering strategy for water electro-oxidation iridate catalysts via coordination distortion. *Chem. Commun.* **55**, 5801–5804 (2019).
 22. Lyons, M. E. G. & Floquet, S. Mechanism of oxygen reactions at porous oxide electrodes. Part 2—Oxygen evolution at RuO₂, IrO₂ and Ir_xRu_{1-x}O₂ electrodes in aqueous acid and alkaline solution. *Phys. Chem. Chem. Phys.* **13**, 5314 (2011).
 23. Murawski, J. *et al.* Benchmarking Stability of Iridium Oxide in Acidic Media under Oxygen Evolution Conditions: A Review: Part I: Probing degradation of iridium-based oxide catalysts. *Johnson Matthey Technology Review* **68**, 121–146 (2024).
 24. Geiger, S. *et al.* The stability number as a metric for electrocatalyst stability benchmarking. *Nat Catal* **1**, 508–515 (2018).
 25. Zhang, D. *et al.* Optimal Electrocatalyst Design Strategies for Acidic Oxygen Evolution. *Advanced Science* **11**, 2401975 (2024).
 26. Moriau, L. *et al.* Ir metal nanoparticles and IrO₂ for acidic oxygen evolution reaction: Insight from Raman spectroscopy. *Sustainable Materials and Technologies* **40**, e00901 (2024).
 27. Muthee, D. K. & Dejene, B. F. Effect of annealing temperature on structural, optical, and photocatalytic properties of titanium dioxide nanoparticles. *Heliyon* **7**, e07269 (2021).
 28. Johnson, A. Electrochemical Surface Area (ECSA) Evaluation in Electrocatalysis: Principles, Measurement Techniques, and Future Perspectives. *J. Eng. Ind. Res.* **6**, (2025).
 29. Voiry, D. *et al.* Best Practices for Reporting Electrocatalytic Performance of Nanomaterials. *ACS Nano* **12**, 9635–9638 (2018).
 30. Cignoni, P. *et al.* Validating Electrochemical Active Surface Area Determination of Nanostructured Electrodes: Surface Oxide Reduction on AuPd Nanoparticles. *J. Electrochem. Soc.* **170**, 116505 (2023).

Evaluation of PEGDA-based microgels as a platform for tattoo biosensing

Callum Yuan Tung and Gabriel Pradadeville

Department of Chemical Engineering, Imperial College London

Abstract

Microgels are a promising progression in tattoo biosensor technology, allowing for immobilisation of complex molecules that improves the viability of long-term usage in medical diagnostics. By developing PEGDA-700 microgel polymers infused with a methylene blue solution, the feasibility of said microgel with colourimetric and fluorescent biosensors could be assessed via a microplate reader with samples at different pHs and a reversible redox reduction of methylene blue. Comparisons between pre photopolymerisation and post photopolymerisation addition of methylene blue to the PEGDA, and the swelling nature of the microgels in different solvents were made and considered. By tattooing the microgels into chicken skin and confirming activity of the methylene blue post-tattooing, this study establishes usage of microgels in tattoo biosensors

Keywords: microgel, biosensor, tattoo, methylene blue, fluorescent, photopolymerisation

1. Introduction

Biosensors are analytical devices that integrate biological components with a physiochemical detector to measure chemical reactions, producing a signal in response to a specific analyte.[1] They have been increasingly important in modern technology and have been applied in many fields ranging from food industries to marine sectors.[2] Owing to their high selectivity, along with rapid and accurate detection capabilities, the use of biosensors in the medical field has been expanded significantly in recent years.

Biosensors have been integrated into wearables that rely on external patches or are embedded in smartwatches to provide real time monitoring of biochemical signals.[3] Most prominent applications of these wearable biosensors are in the form of continuous glucose monitors, where patients with diabetes mellitus can monitor their glucose levels at any time without the need for frequent finger-prick tests.[4], [5]. However, despite their success, most current wearable biosensors suffer from challenges such as poor skin adhesion, limited operational lifetime and general discomfort

during prolonged wear [6] Thus, there is still a need for development of more robust and long-lasting biosensing platforms.

Body modification in the form of tattooing has been around for at least 5000 years.[7]. In the UK, 23% of adults have at least one tattoo and 13% have more than one.[8] Combining aesthetic appeal with the potential for self-diagnosis, tattoos have emerged as a promising platform for integrating biosensing technology [9]. The tattooing process involves puncturing the outer dermis layer of the skin with solid needles, then inserting pigments into the skin where it is localised for extended periods. By replacing traditional tattoo inks with biosensors capable of detecting biochemical changes, the tattoos could offer minimally invasive alternatives capable of long-term usage. Yet issues such as biocompatibility and reversibility are still major barriers that must be addressed before such biosensors can be translated into practical use. One approach to overcome these limitations is by using microgels to encapsulate the biosensors prior to injection.

Microgels are cross-linked polymer particles with characteristics of colloids, macromolecules and surfactants.[10] The ability to swell in response to external stimuli is one of their most defining features. By controlling the swelling of the microgels, they could be tailored for specific uses such as drug delivery [11] or tattoo biosensors where only specific analytes of specific dimensions are allowed to diffuse into the microgels. Derived from toxic free components, microgels are highly biocompatible [12]. By acting as a protective barrier, they can encapsulate biosensing dyes that may not be inherently biocompatible, allowing for a wider range of sensing materials to be explored safely.

The aim of this research is to evaluate the feasibility of PEGDA 700 based microgels as a tattoo biosensing platform. By synthesising the microgels and encapsulating methylene blue dye within them, this study investigates how the microgels influence the dye's fluorescent and colourimetric sensing capabilities.

2. Background

2.1 Microgels

Microgels can be effectively used to encapsulate biosensors; this is particularly important for enzymes, oligomers, and fluorescent dyes. Microgels are vital since they keep the biosensor protected and stable, reducing leaching and improving the stability of the chemical, whilst still allowing the biomarker to diffuse in and change the local environment in the gel. For instance, a paper by Mathew E Allen, et al, outlines the development and application of 'Engineered bacteria encapsulated into micron-scale hydrogel beads prepared through scalable microfluidics'. In this report, they emphasise the importance of encapsulation to enhance functionality, safety, stability, protection, and precise localisation[13].

For the purpose of chemical organic biosensors in tattoos, rather than engineered bacteria, acrylate-based polymers were chosen due to their transparency, colour variation, robust mechanical properties and elasticity.

This is preferred to other synthetic polymers due to their biocompatibility and chemical versatility [14]. In particular, PEG-based microgels were developed since PEG-based hydrogels are widely used as drug delivery and tissue scaffolding materials, this ensures the cytotoxicity of the microgels are negligible.

To create the microgels, poly (ethylene glycol) diacrylate (PEGDA 700), a derivative of polyethylene glycol, is polymerised. A photo initiator is required to trigger photopolymerisation, creating reactive species when exposed to UV light. 2-hydroxy-2-methylpropiophenone was chosen as the photo initiator due to its minimal cytotoxicity at low concentration and its inert nature with the biosensor, methylene blue. (Depending on the biosensors' related biomarker, alternative initiators may be preferable.) The full photopolymerisation mechanism is detailed in section 2.2. The biosensor dye can be added before or after the photopolymerisation and the resulting microgels from these differing approaches are detailed in the report.

2.2 Photopolymerisation mechanism

The microgel formation mechanism is a free-radical photopolymerisation initiated by light which consists of three major steps: initiation, propagation and termination.

Initiation is caused by UV light exciting the 2-hydroxy-2-methylpropiophenone, breaking a C-C bond and forming two free radicals. These attack the double bond of the acrylate group on PEGDA to form a new PEGDA radical. During propagation, the newly formed radical attacks the acrylate group of another PEGDA molecule, continuing to form a polymer chain. Finally, termination occurs when two radicals combine and the chain stops growing. Exact mechanism for this is shown in Figure 2.2.1, a, b and c correspond to initiation, propagation and termination.

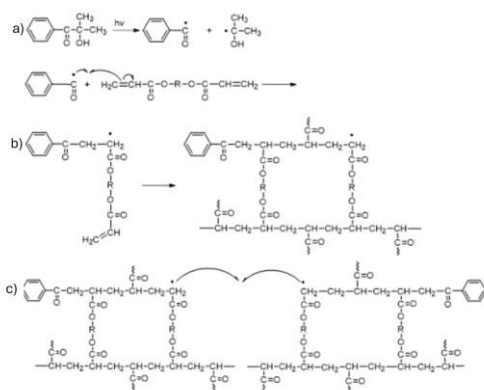


Figure 2.2.1: Free radical photopolymerisation of PEGDA-700

2.3 Methylene blue biosensor

Methylthionium chloride, commonly known as methylene blue, is a synthetic dye first created in 1876. It is one of the first synthetic drugs to ever be used in medicine and has been a cure for malaria for over a century, and in recent years it has been investigated for neurodegenerative disorders [15]. Methylene blue's rich history in medicinal practice proves it to be safe at lower concentrations and viable for tattooing. This along with it being common, cheap and both fluorescently and colourimetrically active made it ideal for testing the PEGDA microgels.

Methylene blue is redox-active phenothiazine dye [16]. This means it undergoes a reversible redox reaction with a distinct colour change, allowing to easily test the viability and reversibility of the microgel with colourimetric biosensors by using a reducing agent. Issues with methylene blue as a colourimetric biosensor come from mediator leaching, since the dye is immobilised in the microgel this is avoided.

In addition, methylene blue is fluorescent with an excitation energy of 665 nm and a peak at 685-688 nm. The fluorescence emission intensity varies with pH, making it viable as a dye to test the microgels' effectiveness with fluorescent biosensors. Therein methylene blue allows to test the microgels' effectiveness both colourimetrically and fluorescently.

2.4 Advances in tattoo biosensing with microgels

Whilst tattoo biosensors for pH and reducing agents have been widely researched, the usage of microgels for more specialised biosensors that indicate for more complex biomarkers is a possible breakthrough for diagnostics in the medical field.

For instance, an immobilised urease and pH indicator mixture could colourimetrically indicate for high concentrations of urea in blood, allowing for early diagnostics of dehydration and kidney dysfunction.

Furthermore, finely designed fluorescent aptamers that bind to cancer-related biomarkers such as cytidine or cardiac disease biomarker Inosine [17] would need immobilisation to avoid leaching and any possible cell damage. Hence microgel technology would be vital to the effectiveness of these biosensors.

3. Practical Methodology

3.1 Materials

All chemicals used were analytical grade. Poly (ethylene glycol) diacrylate (PEGDA 700), sorbitan monooleate (Span 80), 2-hydroxy-2-methylpropiophenone (HMPP), white mineral oil, methylene blue, n-hexane, deionised water.

3.2 Preparation of microgels

Two methods of microgel preparation were performed, pre photopolymerisation and post photopolymerisation. To prepare the pre photopolymerisation microgels, 10mg of methylene blue was weighed with a measuring scale. In this instance, weighing paper was used instead of a weighing boat to minimise powder loss and to reduce static cling. The dye was transferred to a 10mL stoppered volumetric flask and a few drops of deionised water were added to facilitate dissolution. Once the dye had fully dissolved, 1mL of PEGDA 700 was first added into the flask before adding 0.2mL of the photo initiator HMPP. The solution was then made up to 10mL with deionised water before closing the flask with a stopper and shaken to ensure the chemicals were well mixed.

In the meantime, 50mL of white mineral oil was

measured and added to a single neck rounded bottom flask and fixed over a magnetic stirrer with a clamp and stand. A 25 x 10 mm oval shaped magnetic spinner was used, and the stirrer was set to the maximum RPM. 1mL of the emulsifier Span 80 was added to the flask, before adding the dye mixture. It was left to stir for 5 minutes, before directing the flask with a UV lamp for another 5 minutes to trigger the reaction.

For post photopolymerisation microgels, a very similar method was adopted. 1mL of PEGDA 700 and 0.2ml of HMPP were mixed into the stoppered flask without the dye, before making up the solution to 10mL. This solution was added into the mineral oil to stir for 5 minutes, before being subjected to UV light for another 5 minutes. 20 mg of methylene blue dye was measured and dissolved in 10mL of deionised water in another flask and poured into to the mineral oil mixture and mixed for 10 minutes. The microgels were transferred to glass chemical storage bottles and clearly labelled before placing them in the fridge for storage.

3.3 Sample Preparation for Analysis

To prepare the microgels for analysis, 1mL of the microgel mixture was pipetted into each of 6 plastic centrifuge tubes and capped. The samples were placed into a centrifuge and balance was ensured by pairing tubes opposite to each other. The centrifuge was first set to 8000 rpm for 10 minutes to separate the mineral oil from the microgels. After centrifugation, two layers were visible, with the oil forming a top phase and the microgels sitting at the bottom. The oil was carefully decanted, ensuring no microgels were accidentally removed. Next, 1mL of hexane was added into each tube and the samples were centrifuged again at 2500 rpm for another 10 minutes. Afterwards, the solvent phase that formed at the top was removed, before repeating the previous step once more to ensure all the oil was extracted and no more impurities remained.

4. Results & Discussion

4.1 Post and pre photopolymerisation

Addition of the biosensor can be carried out at two distinct stages of microgel creation, before or after photopolymerisation. To fully understand all options available, both methods were used and tested. Firstly, the microscope imaging of post and pre were compared. This can be seen in Figures 4.1.1 and 4.1.2:

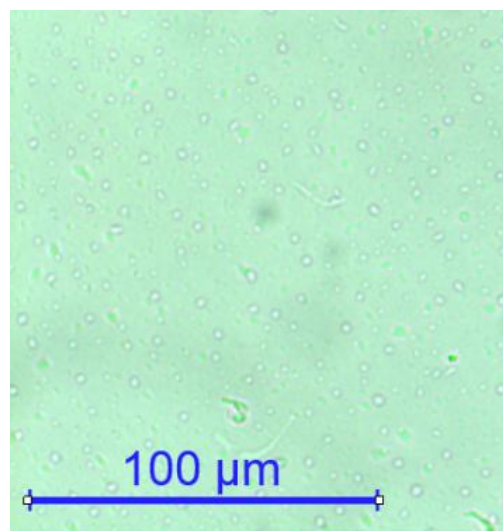


Figure 4.1.1: Microscope imaging for pre photopolymerisation

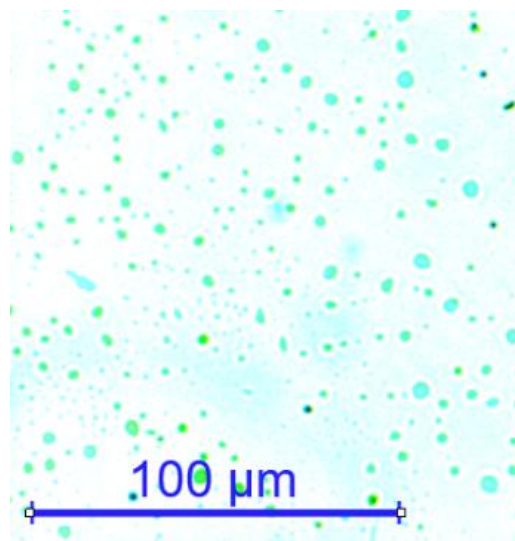


Figure 4.1.2: Microscope imaging for post photopolymerisation.

Using the more clearly visible beads and the microscope scale, the average bead diameter was 3.043 μm with a standard deviation of 0.600 μm for pre and 3.967 μm with a standard deviation of 1.207 μm for post. Due to the smaller beads from pre photopolymerisation and the

methylene blue not requiring to diffuse into the beads compared to post, the concentration of methylene blue in pre photopolymerisation microgels are higher, thus explaining the stronger colour intensity and higher fluorescent emission from those beads, as shown in Figure 4.1.3. Meanwhile, the average volume of methylene blue encapsulated is larger for post photopolymerisation. If you require larger beads on average, post photopolymerisation is preferred. If you want a more responsive biosensor, smaller beads are preferred for the smaller surface area to volume ratio; this improves diffusion into the beads.

Figure 4.1.3 shows the fluorescent emission spectrums of the microgels for both methods:

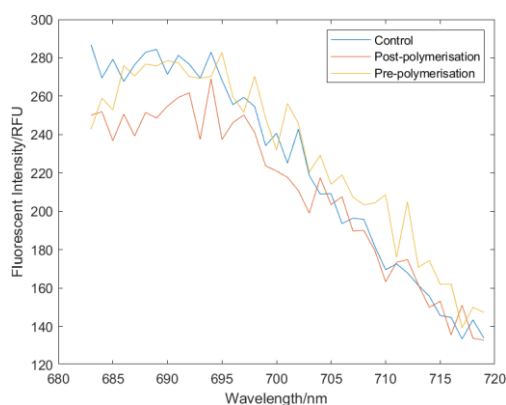


Figure 4.1.3 Fluorescent emission of methylene blue when added post and pre photopolymerisation of the microgel.

From Figure 4.1.3, both microgels have a characteristic peak at 685-688nm, similar to the control solution, which is a solution of methylene blue dissolved in deionised water. This shows that methylene blue fluoresces as expected when immobilised in PEGDA microgel. Furthermore, pre photopolymerisation has a slightly higher peak fluorescent intensity; this is because of the higher concentration of dye encapsulated in the microgel. The relationship between concentration of dye and fluorescent intensity is explored more thoroughly in section 4.4 of the report.

4.2 Absorption spectrum

To further prove that the methylene blue had fully diffused into the microgel for both methods, an absorption spectrum was taken and compared to known

data. The absorption spectrum for methylene blue microgels, both post and pre polymerisation is given below:

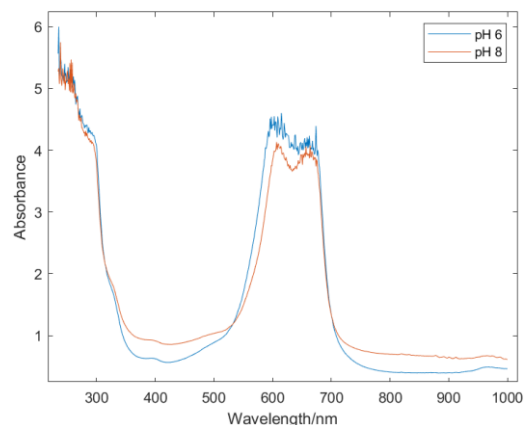


Figure 4.2.1 Absorption spectrum for post photopolymerization

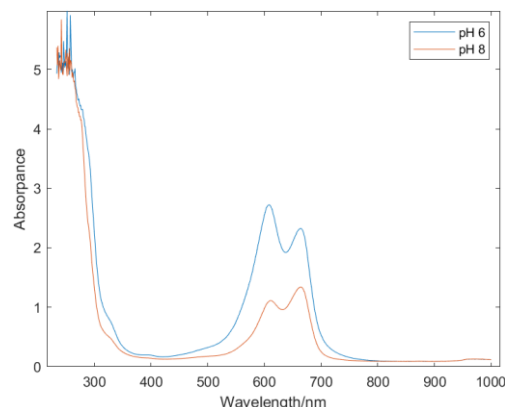


Figure 4.2.2 Absorption spectrum for pre photopolymerisation

Both absorption spectra have the expected peak at 665 nm and a shoulder peak at 615 nm, this neighbouring peak corresponds to the dimmer double protonated methylene blue species [18], this explains the higher amplitude at lower pH as more of the methylene is protonated. The lower, larger peak is for the PEGDA 700 which tends to have a peak around 200-230 nm. This matches the peak in both spectra.

4.3 Fluorescent emission at varying pHs

As mentioned in section 2.3, methylene blue is a fluorescent emitter with an excitation wavelength of 665 nm and peak fluorescent emission at 685-688nm. Whilst the peak does not shift due to changes in pH, the fluorescent intensity does. This is caused by doubly protonated or partially reduced methylene blue being more prevalent at lower pHs, these either do not fluoresce

or have lower quantum yields. Hence fluorescent intensity lowers. To test for this, tris-buffered pH solutions of 6,7 and 8 were synthesised. The effect of pH on intensity is shown in Figure 4.3.1 for pre photopolymerisation and Figure 4.3.2 for methylene blue addition post photopolymerisation:

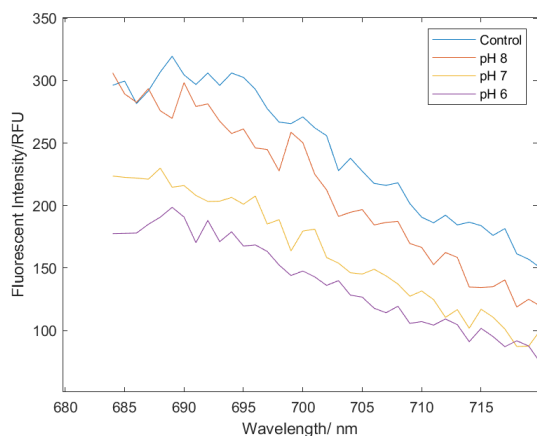


Figure 4.3.1 Fluorescent emission at varying pHs for pre Photopolymerisation addition of methylene blue

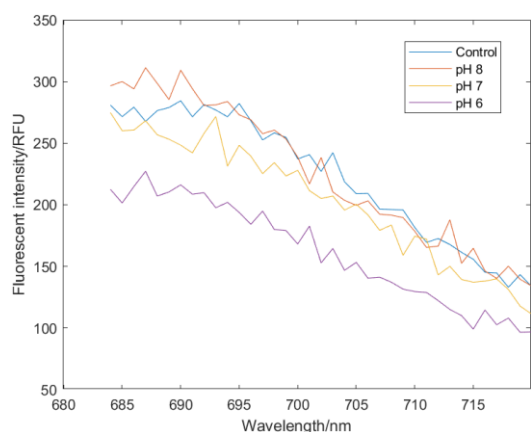


Figure 4.3.2 Fluorescent emission at varying pHs for post photopolymerisation addition of Methylene blue.

In Figure 4.3.1, the peak fluorescent intensity is 306.1, 230.0 and 198.6 (± 0.05) RFU for pH 8,7 and 6 respectively, with a total drop of 107.6 over the 3-pH range. With the overall trend being lower intensity at equivalent wavelengths. Similarly, in Figure 4.3.2, peak fluorescent intensity is 311.2, 268.2, 227.2 (± 0.05) RFU for pH 8,7 and 6 respectively, with a total drop of intensity of 82 RFU and a trend of lower intensity at equivalent wavelengths. Post photopolymerisation having a smaller drop in fluorescent intensity is due to it having a higher volume of methylene blue in each microgel and smaller surface area to volume

ratio; this means a smaller proportion of the dye is protonated to its less fluorescent form and the intensity is affected less.

From this one can confirm that the fluorescent activity of a biosensor encapsulated in the PEGDA microgels is affected by its local exterior environment, hence, its usage in tattooing would be useful for detecting biomarkers so far as the effectiveness of the biosensor.

4.4 Fluorescent emission at different concentrations of methylene blue microgels.

Concentration of the biosensor microgel is an important consideration when designing the tattoo ink. To see how concentration affected fluorescent intensity, four solutions of microgel at 500, 200, 100 and 50g/L were prepared.

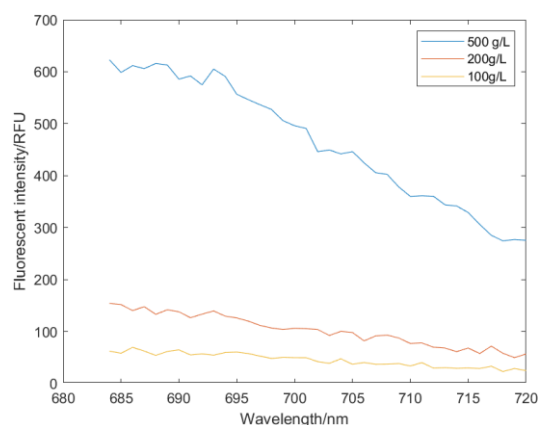


Figure 4.4.1 Fluorescent intensity of methylene blue microgel at varying concentrations

50g/L was emitted due to its fluorescent emission being too minimal to be detected accurately. From Figure 4.4.1 the fluorescent intensity of a sample is strongly correlated with the concentration of methylene blue microgel. With the peak emission intensity being 622.9, 154.1 and 69.51 RFU (± 0.05) for 500, 200 and 100g /L respectively. This matches expectations as there is less fluorescent quantum yield for solutions with a smaller proportion of methylene blue. When designing dye concentration in tattoo, it is important to consider the intensity of fluorescence emission to ensure easy detection and diagnostics whilst also balancing the safety of the wearer and the colouration of the ink.

4.5 Reduction of methylene blue

When added to a reducing agent, methylene blue undergoes a reversible redox reaction to leucomethylene, this reaction is shown in figure 4.5.1 with methylene blue above and leucomethylene below:

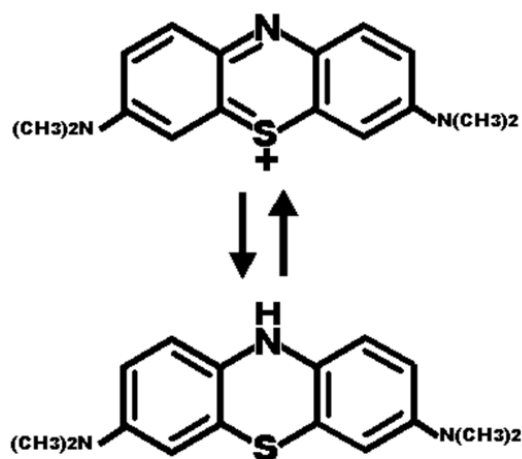


Figure 4.5.1: Reversible redox reaction of methylene blue.

This is a colourless solution providing a colour change. This allows to test for colorimetric diagnostic and reversibility whilst encapsulated in the acrylate microgel. To test this, the ‘blue bottle’ experiment was used. By adding the methylene blue microgel to a solution of glucose, mixing and letting it settle will turn the solution colourless. Shaking the solution again will re-introduce oxygen, causing it to oxidise and return to a blue colouration. These stages are shown in Figure 4.5.2:

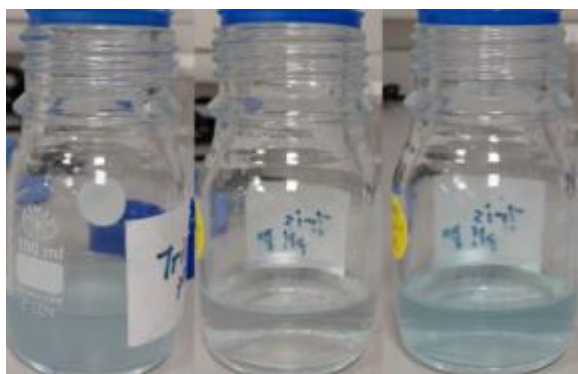


Figure 4.5.2: 3 stages of the ‘blue bottle’ experiment

First stage is when methylene blue is initially added to the glucose solution. Second stage is mixing and letting it settle. Third is mixing again and oxidising the solution. There is an observable change in colour at each stage; from blue to colourless to blue once again. Therefore, the microgels

clearly work with a colourimetrically and reversibly. This reversibility ensures that the biosensors can continuously react to any changes in the ISF composition. This is essential for effective diagnostics.

4.6 Swelling properties.

The swelling of microgels is an important characteristic to understand. The ability to tune the dimensions of the microgels can be very valuable. In the case of tattoo biosensors, by controlling the size of the microgels specific analytes can be selectively allowed to diffuse into the microgel and be detected by the biosensor with minimal interference from other biomarkers.

To understand the swelling behaviour of the microgels when in solution, they were suspended in three solvents: ethanol, phosphate-buffered saline and water. These solutions were left for a day, and the particle sizes were characterised using the Anton Paar PSA 1190, where particle size distribution curves below were obtained.

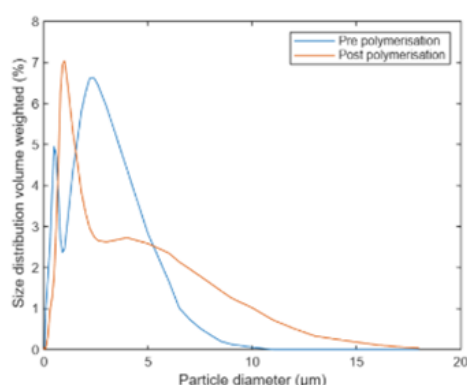


Figure 4.6.1 Microgel swelling behaviour in water.

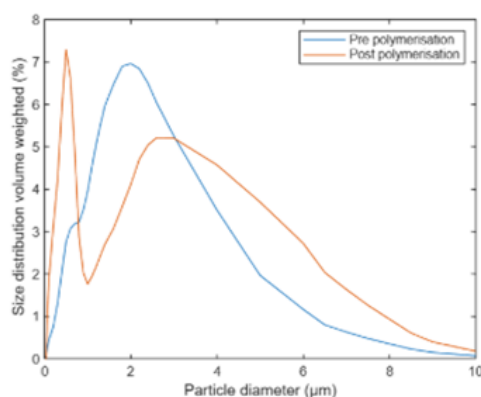


Figure 4.6.2 Microgel swelling behaviour in phosphate buffered saline.

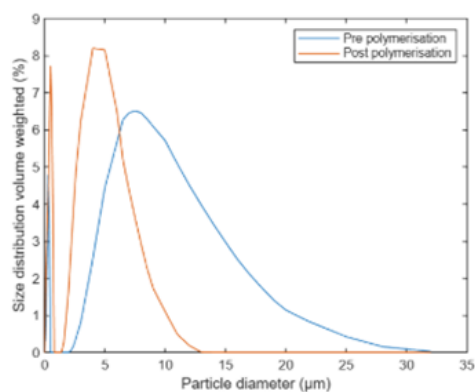


Figure 4.6.3 Microgel swelling behaviour in ethanol.

The swelling behaviour followed similar trends across all three solvents, with pre photopolymerisation microgels more susceptible to swelling compared to their post photopolymerisation counterparts. This can be explained as the post photopolymerisation microgels contained greater volumes of methylene blue. This forms additional crosslinks within the microgels, therefore less susceptible to swelling.

Microgels suspended in ethanol exhibited the largest mean diameters, measuring $5.41\mu\text{m}$ (pre) and $2.41\mu\text{m}$ (post). In contrast, the size distributions in water were the least variable out of the three, with standard deviations of $1.40\mu\text{m}$ (pre) and $2.03\mu\text{m}$ (post) contributing to a narrower spread in particle sizes. These results suggest that water provides the most controlled swelling behaviour, while ethanol is optimal for promoting the formation of the largest swollen microgels. Microgel size distribution in phosphate buffered saline was similar to that of water but with slightly smaller post photopolymerization microgels. Phosphate buffered saline is commonly used in biological research and has similarities to the interstitial fluid. This controlled swelling behaviour suggests that the microgels are likely to behave similarly in the interstitial fluid, with lower risks of over swelling and dye leakage when used as tattoo biosensors.

However, the swelling behaviour of the microgels is not purely governed by the solvent, but is also influenced by other factors such as temperature, pH and light. To fully understand this phenomenon, these variables should be investigated and characterised. Doing so can allow for

more precise control over the swelling of microgels and enable tailoring the diameters for specific applications in the future.

4.7 Tattooing functionality.

To test as a tattoo, samples of the pre and post photopolymerisation methylene blue microgels were injected into chicken skin. Chicken skin was used as it best resembles human skin. Fat was removed from the chicken skin beforehand to minimise interference and for a less greasy surface. Figure 4.7.1 shows the tattooed dye, with post photopolymerization methylene blue microgel above and pre photopolymerisation methylene blue microgel below.



Figure 4.7.1 Microgels tattooed in chicken skin.

By submerging sections of the skin in pH solutions, effectiveness of the methylene blue microgel as a fluorescent biosensor whilst tattooed can be confirmed using the microplate reader:

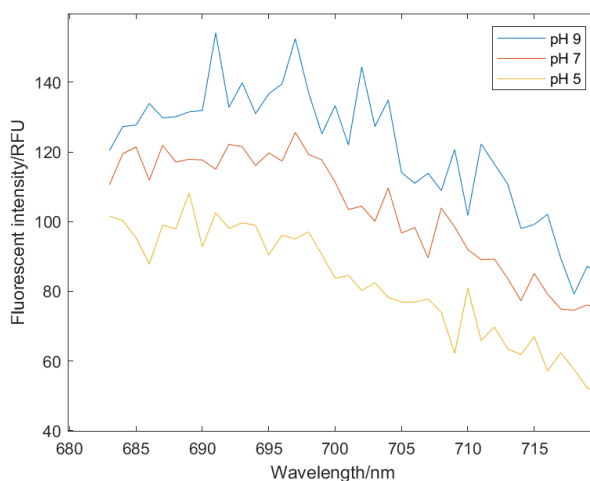


Figure 4.7.2 Fluorescent emissions of methylene blue of varying pH on the chicken skin for post photopolymerisation

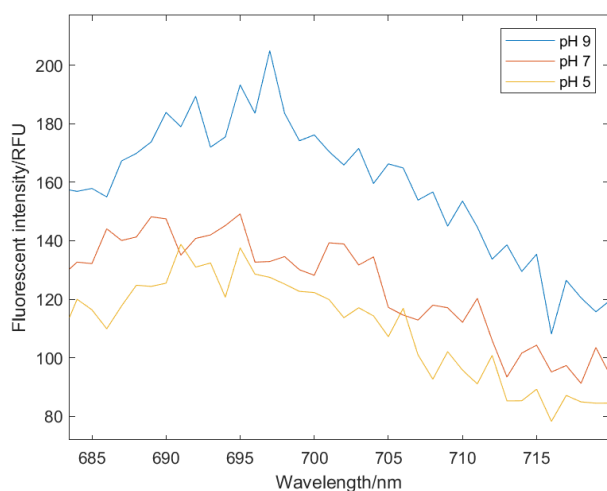


Figure 4.7.3 Fluorescent emissions of methylene blue of varying pH on the chicken skin for pre photopolymerisation

As before, the fluorescent emission peak intensity decreases with pH, peaking at 154.2, 122.2 and 108.2 RFU (± 0.05) for pH 9, 7 and 5 respectively for post-photopolymerisation and 193.3, 148.2 and 138.3 for pre-photopolymerisation. Due to it being tattooed in the skin, the peak is broader, of lower intensity, and a larger range of pHs is required to get distinct curves with no overlap; therefore, overall sensitivity of the biosensor is decreased. The tattoo ink did not fade overnight or when exposed to UV light, this tells us that it's stable and doesn't leach away. All together this confirms the viability of microgels in tattooing as a diagnostic tool but drop in intensity needs to be considered when designing the dye solution.

5. Conclusion

In this work PEGDA based microgels were synthesised and methylene blue was successfully encapsulated within them. The microgels' influence on the dye's performance was evaluated, and their effect on the dye's biosensing capabilities and reversibility were tested. Results from the microplate reader confirmed the presence of the dye within the microgels, and the dye's sensitivity was minimally affected. The microgels synthesised using the two different methods had similar physiochemical properties, but clear differences were observed. Post photopolymerisation microgels were larger in general and more stable in size, while pre photopolymerisation microgels were more susceptible to swelling. When tattooed onto chicken skin, although the biosensor's functionality wasn't affected, the

sensitivity however was lowered. Further studies must be conducted to determine strategies to minimise sensitivity loss when tattooed. Additionally, the microgels could be subjected to different environmental conditions such as by changing temperatures and humidity, or to simulate sweating to assess the tattoo's stability and ensure long term functionality.

Swelling of microgels is of particular interest in research, and while pre photopolymerisation microgels demonstrated better swelling in solvent, this behaviour is also affected by other external stimuli and needs to be investigated further. Alternative biosensors could also be explored to test the compatibility of these microgels, and future work may involve tattooing the microgels into a medium that better mimics the interstitial fluid in the dermis layer of the skin for more representative results. Nevertheless, these PEGDA based microgels show considerable promise as a platform for tattoo biosensing.

6. Acknowledgements

The authors express their sincere gratitude to Nina Petrovičová for providing the research framework and original experimental protocols in which this study is built. Her guidance and support, along with that of the Yetisen group, greatly contributed to the overall quality of this work.

7. References

- [1] 'Biosensors for Analytical Purposes | BMG LABTECH'. Accessed: Nov. 02, 2025. [Online]. Available: <https://www.bmglabtech.com/en/blog/biosensors/>
- [2] P. Mehrotra, 'Biosensors and their applications – A review', *J. Oral Biol. Craniofacial Res.*, vol. 6, no. 2, pp. 153–159, 2016, doi: 10.1016/j.jobcr.2015.12.002.
- [3] J. Kim, A. S. Campbell, B. E.-F. de Ávila, and J. Wang, 'Wearable biosensors for healthcare monitoring', *Nat. Biotechnol.*, vol. 37, no. 4, pp. 389–406, Apr. 2019, doi: 10.1038/s41587-019-0045-y.
- [4] 'Wearable/disposable sweat-based glucose monitoring device with multistage transdermal drug delivery module'. Accessed: Nov. 04, 2025. [Online]. Available: <https://www.science.org/doi/10.1126/sciadv.1601314>

- [5] 'Continuous glucose monitoring and hybrid closed loop for diabetes', nhs.uk. Accessed: Nov. 04, 2025. [Online]. Available: <https://www.nhs.uk/tests-and-treatments/cgm-and-hcl-for-diabetes/>
- [6] A. Ahmad Tarar, U. Mohammad, and S. K. Srivastava, 'Wearable Skin Sensors and Their Challenges: A Review of Transdermal, Optical, and Mechanical Sensors', *Biosensors*, vol. 10, no. 6, p. 56, May 2020, doi: 10.3390/bios10060056.
- [7] 'A brief history of tattoos', Wellcome Collection. Accessed: Nov. 09, 2025. [Online]. Available: <https://wellcomecollection.org/stories/a-brief-history-of-tattoos>
- [8] samuelneave, 'Why we get tattoos', Opinium. Accessed: Nov. 04, 2025. [Online]. Available: <https://www.opinium.com/why-we-get-tattoos/>
- [9] A. K. Yetisen *et al.*, 'Dermal Tattoo Biosensors for Colorimetric Metabolite Detection', *Angew. Chem. Int. Ed.*, vol. 58, no. 31, pp. 10506–10513, 2019, doi: 10.1002/anie.201904416.
- [10] M. Karg *et al.*, 'Nanogels and Microgels: From Model Colloids to Applications, Recent Developments, and Future Trends', *Langmuir*, vol. 35, no. 19, pp. 6231–6255, May 2019, doi: 10.1021/acs.langmuir.8b04304.
- [11] H. Omidian and S. Dey Chowdhury, 'Swellable Microneedles in Drug Delivery and Diagnostics', *Pharmaceuticals*, vol. 17, no. 6, p. 791, June 2024, doi: 10.3390/ph17060791.
- [12] Md. S. Ahmed, S. Yun, H.-Y. Kim, S. Ko, M. Islam, and K.-W. Nam, 'Hydrogels and Microgels: Driving Revolutionary Innovations in Targeted Drug Delivery, Strengthening Infection Management, and Advancing Tissue Repair and Regeneration', *Gels*, vol. 11, no. 3, p. 179, Mar. 2025, doi: 10.3390/gels11030179.
- [13] M. E. Allen *et al.*, 'Engineered Bacteria as Living Biosensors in Dermal Tattoos', *Adv. Sci.*, vol. 11, no. 30, p. 2309509, 2024, doi: 10.1002/advs.202309509.
- [14] K. Krutkramelis, B. Xia, and J. Oakey, 'Monodisperse Polyethylene Glycol Diacrylate Hydrogel Microsphere Formation by Oxygen-Controlled Photopolymerization in a Microfluidic Device', *Lab. Chip*, vol. 16, no. 8, pp. 1457–1465, Apr. 2016, doi: 10.1039/c6lc00254d.
- [15] R. H. Howland, 'Methylene Blue: The Long and Winding Road from Stain to Brain: Part 1', *J. Psychosoc. Nurs. Ment. Health Serv.*, vol. 54, no. 9, pp. 21–24, Sept. 2016, doi: 10.3928/02793695-20160818-01.
- [16] L. Fu, C.-T. Lin, and K.-M. Hassan, 'Methylene blue in electrochemical (Bio)sensing: Historical evolution, mechanistic insights, and emerging applications—A review', *Anal. Chim. Acta*, vol. 1377, p. 344468, Dec. 2025, doi: 10.1016/j.aca.2025.344468.
- [17] P. P. Samant *et al.*, 'Sampling interstitial fluid from human skin using a microneedle patch', *Sci. Transl. Med.*, vol. 12, no. 571, p. eaaw0285, Nov. 2020, doi: 10.1126/scitranslmed.aaw0285.
- [18] J.-C. Bollinger, E. C. Lima, L. Mouni, S. Salvestrini, and H. N. Tran, 'Molecular properties of methylene blue, a common probe in sorption and degradation studies: a review', *Environ. Chem. Lett.*, vol. 23, no. 5, pp. 1403–1424, Oct. 2025, doi: 10.1007/s10311-025-01856-1.

Multi-solvent computer-aided molecular and process design framework to predict VLE properties

Isabella McGrath and Lucia Gibilaro

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Carbon capture technologies must be optimised to reduce carbon dioxide (CO₂) greenhouse gas emissions, but current amine solvents require high energy regeneration costs. Rigorous thermodynamic models to predict vapour-liquid equilibrium (VLE) properties impose substantial computational burdens that limit large-scale solvent screening. This work extended the single-solvent surrogate approach of Baakes *et al.*³ to a multi-solvent computer-aided molecular and process design framework using an artificial neural network (ANN) surrogate to predict VLE properties from molecular group count (GC) and process condition inputs. 52 diverse amine solvents were selected from a Mixed Integer Linear Programming generated and PubChem combined dataset using k-means clustering and D-optimal sampling. Two ANN models were mapped 19 inputs (including 13 GC descriptors and process conditions) to 17 VLE outputs across 4,096 and 16,384 datapoints per amine. Model 2 achieved an overall MAE of 0.0179, with four outputs successfully meeting all defined accuracy thresholds, including vapour CO₂ composition ($R^2 > 0.97$). Despite this, substantial failures occurred in compositional predictions, all water predictions and the vapour-to-liquid ratio (β). MEA out-of-sample validation yielded higher overall error (MAE = 0.032) but improved water mole number predictions. Overfitting analysis showed minimal training-test performance gap, suggesting core limitations are insufficient sample size (52 amines for 19-dimensional input space) and inadequate GC representation. Results demonstrate that multi-solvent CAMPD is feasible for pre-screening amine solvents for carbon capture absorption, but requires increased sample size, consideration of molecular descriptors, and hyperparameter optimisation before use in process design.

1. Introduction and Background

Due to the urgent environmental crisis directly resulting from anthropogenic activity, effective methods to reduce greenhouse gas emissions, particularly carbon dioxide (CO₂), must be found to mitigate climate change and global warming. In 2024, global energy-related CO₂ emissions rose to a record high of 37.8 billion tonnes, a 0.8% increase from 2023, contributing to unprecedented atmospheric CO₂ concentrations of 422.5 ppm.¹¹ The resulting movement towards more sustainable industrial practices means that optimising carbon capture technologies has become a major area of interest in the scientific community. Carbon capture refers to technologies which separate CO₂ from exhaust gas streams from industrial processes, for instance, in natural gas processing, power generation, and cement and steel production. The traditional approach to carbon capture, used for over 70 years, is amine-based scrubbing.²

This method uses chemical absorption, in which CO₂ is captured via a reversible acid-base reaction with an amine-based solvent, typically an aqueous alkanolamine such as monoethanolamine (MEA), forming a bicarbonate and carbamate species.³ The solvent is subsequently thermally regenerated, desorbing concentrated CO₂. Advantages of these systems include their extensive industrial experience, their capacity to achieve high CO₂ removal levels even at low CO₂ partial pressures/concentrations, and their ability to be retrofitted to existing flue gas streams and plants relatively easily.⁴ However, a significant issue with these processes is the substantial energy demand for solvent regeneration, which increases

operational costs, reduces the likelihood of large-scale industrial deployment, and limits their environmental benefits due to additional fuel requirements.⁵

Benchmark solvents currently used for such chemical absorption processes include MEA, diethanolamine (DEA), and methyl diethanolamine (MDEA); however, it is evident that solvent choice optimisation is essential to reduce energy demands and improve the performance of carbon capture technologies. Due to the large number of potential amines, a systematic and computational approach is required to explore the chemical design space effectively, with recent research focusing on integrated molecular and process methodologies⁶

Computer-aided molecular and process design (CAMPD) is a promising area of focus that provides a structured method for assessing a wide range of molecular structures based on physicochemical properties and process performance metrics.⁷ The framework uses predictive thermodynamic models, in our case, the SAFT- γ Mie group-contribution equation of state (EoS), which predicts the phase-equilibrium properties for a wide range of solvent molecules. Group contribution methods, which assume that molecular properties can be predicted from the summed contributions of their constituent functional groups, are used due to their simplicity, computational efficiency, and high transferability.⁶ Despite the success in finding more effective solvents than those found with sequential molecular design approaches, an integrated CAMPD solution has the current limitation of a high computational cost, due to the nonlinear interactions between physical properties and

process performance. This is highlighted by Lee *et al.*⁷ who developed a fully integrated CAMPD framework for amine-based CO₂ capture using the SAFT- γ Mie EoS.

To overcome this substantial computational burden, surrogate models can be utilised; these are less demanding, mathematically simple models that approximate the relationship between inputs and outputs and are becoming increasingly popular within the chemical process engineering field due to their applications in process optimisation.⁸ The recent growth in machine learning (ML) accessibility has also enhanced their applicability, allowing for the development of flexible, data-driven ML surrogate models that can replicate complex, non-linear relationships with higher accuracy, lower execution speed, and lower computational expense.⁹ Amar *et al.* (2019)¹⁰ demonstrated this by employing a Gaussian process surrogate model combined with descriptors to predict solvent effects in asymmetric catalysis, enabling efficient exploration of a large solvent library and identifying promising solvents. While Gaussian process models have shown some success in solvent screening, artificial neural networks (ANNs), which are employed in this project, are a more commonly used ML algorithm that performs well for high-dimensional, complex, non-linear systems.⁹ ANNs have proven to be effectively applied in surrogate-model-based optimisation studies, and Zhang *et al.* (2021) illustrated this through the development of an ANN-based surrogate to reproduce CO₂ solubility behaviour within the molecular and process design of ionic-liquid absorption systems.¹¹ Global optimisation for both solvent and process parameters was achieved with substantially reduced computational demand. Notable prior work on amine-based carbon capture includes Baakes *et al.*³, which demonstrated the feasibility of using a surrogate ANN model to accurately replicate the vapour-liquid equilibrium (VLE) behaviour of aqueous MEA predicted by the SAFT- γ Mie EoS, significantly reducing computational time and effectively exploring the design space compared with the rigorous thermodynamic process model.³

Our work builds on this study, extending the approach from a single-solvent CAPD into a multi-solvent comprehensive CAMPD framework, targeting efficient representation of the amine solvent design space. This project trains an artificial neural network (ANN) surrogate model to reproduce the group contribution SAFT- γ Mie EoS and predict VLE properties from molecular structure and process conditions, thereby reducing computational expense for carbon capture optimisation. This report places particular emphasis on defining the solvent design space, the selection of two- and three-dimensional (-D) molecular descriptors through data-driven analysis, and

evaluating the clustering and D-optimal sampling strategy employed to select the set of solvent molecules used to train, test, and validate the final surrogate model. The computational benefits and predictive accuracy of this approach were evaluated against the traditional, rigorous thermodynamic SAFT- γ Mie model.

2. Methodology

2.1 Solvent Design Space Definition

This section defines the two design spaces from which the final optimal sets of solvent molecules were derived. Both selection spaces consist solely of uncharged, saturated, acyclic amine-based solvents, constructed from a set of thirteen predefined functional groups. These functional groups are CH₂NH₂, CHNH₂, CNH₂, CH₃NH, CH₂NH, CHNH, CH₃N, CH₂N, CH₃, CH₂, CH, C, and OH. Every candidate solvent molecule in each selection space has been encoded using two molecular representations: a canonical SMILES string and a group-count (GC) vector. For a given solvent molecule, its GC vector $\mathbf{n}$ stores thirteen GCs n_i , defined as the number of times each functional group i appears in its structure.

The first selection space was constructed using a mixed-integer linear programming (MILP) model that systematically enumerated possible combinations of the 13 functional groups, subject to a set of structural and physical property constraints. The problem was formulated with 13 integer decision variables ($n_1, n_2, \dots, n_{13}$), each constrained by lower and upper bounds $n_i \in [\text{LB}, \text{UB}]$, listed in Table SI.2. No objective function was specified; instead, a constant (dummy) objective of zero was set so that the problem identified all integer solutions that satisfied the variable bounds, one equality constraint, and seven inequality constraints. The constraints are as follows: (1) each structure was required to comply with the octet rule; (2) the summed count of all functional groups was constrained to at most 12; (3) the number of nitrogen-based functional groups was restricted to a maximum of 2; (4) at least one nitrogen-based functional group had to be present; (5) where a hydroxyl group was present, at least one of the methylene-containing groups (CH₂NH₂, CH₂NH, CH₂N, or CH₂) must be present; (6) an upper bound of 4 CH and C groups was enforced; (7) a lower bound of 308.15 K was set for the flash point of each theoretical molecule; and (8) an upper bound of 300 g/mol for the molecular weight. The flash point and molecular weight of each theoretical solvent molecule were determined using the Constantinou and Gani property-group-contribution method; the individual contributions of each functional group are listed in Table SI.1. Each integer solution corresponded to a theoretical solvent molecule, represented by its GC vector. Following the GC vector generation, a structure-construction step was implemented to ensure each

vector could be converted into a valid molecular graph and to obtain the canonical SMILES string of the resulting valid molecule. This step was also formulated as an MILP problem with a dummy objective function of zero, in which the functional groups that composed each theoretical molecule were assembled subject to a set of bonding constraints. Any theoretical molecule that was not chemically and/or structurally feasible was excluded. Both MILP problems were solved using a Gurobi solver.

The second selection space was derived from the PubChem database. The SMILES representations for multiple amine-based molecules were imported, filtered, and then pre-processed. The filters implemented ensured that the selection space consisted exclusively of saturated acyclic molecules with a maximum molecular weight of 300 g/mol and that could be constructed from the 13 predefined functional groups. The data preprocessing involved ensuring the SMILES strings were in canonical form and generating the GC vectors. The vectors were constructed by converting the canonical SMILES strings into SMARTS representations and then performing substructure matching against the SMARTS representations of each predefined functional group.

2.2 Molecular Descriptor Calculation and Screening

A comprehensive set of molecular descriptors was computed for each candidate amine using the canonical SMILES strings. These include 2D descriptors (topological, electronic, H-bonding, and surface area properties) and 3D descriptors (radius of gyration, asphericity, inertial shape factor, and sphericity index). 2D molecular descriptors were computed using RDKit, while 3D molecular descriptors were generated using the ETKDG algorithm, followed by UFF-based geometry optimisation where possible. Molecular identifiers (SMILES and GC vectors) were retained throughout. These descriptor datasets were then screened to remove redundant and non-informative features and reduce dimensionality. Descriptors exhibiting near-zero variance (standard deviation $< 1 \times 10^{-3}$) were removed, and a correlation-based filter was applied to remove one descriptor from any highly correlated pair (Pearson correlation $|r| > 0.97$). Key interpretable descriptors were prioritised, including MolLogP (lipophilicity), TPSA (topological polar surface area), NumHDonors and NumHAcceptors (hydrogen-bonding capacity), RingCount (ring topology) and MaxPartialCharge (electronic distribution). Only those that met the variance and correlation criteria were retained, alongside the three descriptors with the highest variance remaining following filtering. The identical pipeline was applied to both the

PubChem and the generated amines datasets to ensure methodological consistency.

2.3 Sampling and Diversity Selection

Following descriptor screening, two informative and representative sets of feasible solvent molecules were identified from each solvent selection space. This was conducted using a cluster-based D-optimal design methodology that combined k-means clustering with a Fedorov exchange algorithm within each cluster. All molecular descriptors were initially standardised using robust z-score scaling (RobustScalar). All molecular descriptors were standardised using robust z-score scaling. Molecules with any descriptor exhibiting $Z > 3$ were excluded as outliers. Principal component analysis (PCA) was subsequently performed to reduce the dimensionality of the descriptor space before the k-means clustering. The optimal number of clusters (k) and principal components for each selection space was determined using the Elbow method. For every principal component, a curve of the within-cluster sum of squares (WCSS), against k was generated for $k \in [5, 40]$. PCA dimensions ranged from 2 to the total descriptor number per molecule post screening, p . The Kneedle algorithm (KneeLocator) was applied to identify the ‘elbow’ point on each curve. The PCA dimension that yielded an ‘elbow’ point with the lowest WCSS was selected with its corresponding optimal k value. Outliers in the second principal component (PC2) were removed using an exclusion threshold ($PC2 > 10$) as determined during preliminary data inspection. After removing outliers, k-means clustering was performed on each dataset.

To obtain the initial optimal solvent sets, a Fedorov exchange algorithm was applied to each cluster independently. A target number of ~50 feasible solvent molecules (~25 from each dataset) was set. Based on preliminary feasibility tests, initial targets of 50 generated and 250 PubChem amines were selected to obtain approximately 25 feasible molecules from each dataset. To achieve this, the number of candidate solvent molecules to be selected from each cluster (the quota, q) was set proportional to the number of solvent molecules in each cluster l . For each cluster, five random subsets containing q candidate solvent molecules were generated, and their D-optimality score was obtained.

$$D - score = \log(\det(X_n^T X_n))$$

Where X_n is the information matrix (size $q \times p$) containing the scaled descriptor values for each molecule in the n^{th} selected subset of candidate solvent molecules. Each subset was subsequently optimised by maximising its D-optimality score using the Fedorov exchange algorithm over up to 50 iterations. In each iteration, some selected molecule i , where $i = 1, \dots, q$, was replaced by some unselected molecule j , where $j = 1, \dots, (l -$

q), and the descriptors for molecule i were replaced with those for molecule j in the information matrix to form a new matrix, X_{n+1} . The D-optimality score of this new matrix was computed, and molecule swap was accepted only if the D-score increased. The subset with the largest D-score across the five initialisations was retained; the optimal subsets for each cluster were combined to form the initial optimal sets for each design space. These optimal sets were then screened through a feasibility test. This feasibility test confirmed that vapour-liquid equilibrium (VLE) could be achieved by running a gPROMS simulation and any infeasible candidate solvent molecules were excluded. Descriptor coverage was defined as the ratio of a descriptor’s range within a specific dataset to its total range in the full dataset. The overall coverage for each set was computed as the geometric mean across all descriptors.

2.4 Surrogate Model Development

For each feasible amine, VLE data was generated using gPROMS across typical operating conditions (298-410K, 0.96-3.5 bar) and compositions. We collected both 4,096- and 16,384-datapoint sets for each amine. These two datasets consisted of input variables (temperature, pressure, GC vectors, and gas-feed compositions) and output variables (liquid and vapour mole numbers for H₂O, amine, CO₂ and N₂, the vapour fraction, and liquid- and vapour-phase compositions). Pure-phase points were removed, invalid points were excluded, and inputs and outputs were standardised prior to training. The processed datasets were then randomly split into training (80%), validation (10%) and test (10%) subsets. Two feed-forward neural network regression models (with ELU activation functions and the hyperparameters shown in Table 1) were trained using the training subset to map the 19 inputs to four latent thermodynamic outputs. These latent variables were subsequently reconstructed, together with the unscaled feed compositions, into the full set of 17 VLE outputs using an RMSE loss function, which enforced fundamental physical constraints including mass balance, mole fraction normalisation, and VLE relationships. Training used an Adam optimiser with learning rate scheduling over a fixed number of epochs, and stochastic weight averaging (SWA) was applied during the final portion of training. Hyperparameters were chosen through iterative testing for the smaller model and were adapted accordingly for the larger model.

Model	Hidden Size	Hidden Layers	Epochs
Model 1 (4096 datapoints)	100	8	300
Model 2 (16,384 datapoints)	150	8	250

Table 1: Selected hyperparameters (hidden size, depth, and epochs) for the ANN surrogate models

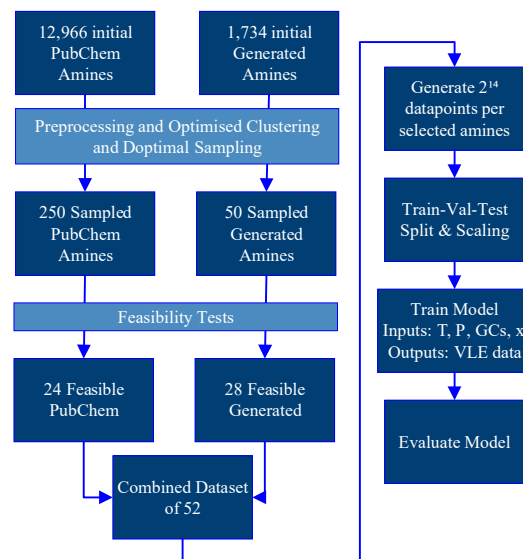


Figure 1: Methodology Workflow for Solvent Selection and Surrogate Model Training and Evaluation

2.5 Model Evaluation

For evaluation, the trained surrogate model architecture was reconstructed, and the model parameters from the most recent, best-performing epoch were loaded. The model was first evaluated on the test set (10% of datapoints from the 52-amine training dataset). Predictions for this set were generated in latent space and converted to the physical VLE outputs using the same reconstruction procedure applied in training. Evaluation metrics comparing the true and predicted values, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and coefficient of determination (R^2), were computed and saved for all 17 outputs. These were compared against defined thresholds shown in Table 2. Parity plots were also generated to provide a visual comparison of model performance. This test set contained data from all 52 amines at different operating conditions, so this evaluation showed the model’s predictive performance across operating conditions rather than its capability to predict VLE properties for new amine structures. To assess this, the model was tested using MEA, the industrial benchmark solvent (not included in the 52-amine sample); 16,384 VLE datapoints across operating conditions were generated for MEA, and the same evaluation metrics (MAE, RMSE, R^2) and parity plots were generated, using identical methodology.

Output	MAE	RMSE	R^2
V-L mole numbers (nx_i , ny_i) and compositions (x_i , y_i)	≤ 0.01	≤ 0.0125	≥ 0.95
Vapour-to-liquid ratio (β)	≤ 0.02	≤ 0.025	≥ 0.95

Table 2: Defined accuracy criteria (MAE, RMSE and R^2) used to benchmark surrogate-model performance for vapour-liquid compositions and vapour-to-liquid ratio.

Model 2 was also evaluated on the training dataset to assess overfitting

3. Results

3.1 Chemical-Space Definition and Dataset Preprocessing

In the first solvent selection space, the MILP formulation generated a total of 4,799 group count (GC) vectors. Following the structure-construction step, this set was reduced to 3,069 chemically valid molecules that satisfied all bonding requirements and could be parsed to valid molecular graphs. After removing GC vectors producing duplicate canonical SMILES strings, a final number of 1734 unique structures were identified. The final PubChem amine dataset and the generated amine dataset, defined as the selection spaces, consisted of 12,966 and 1,734 amines, respectively. These two design spaces contained 117 common candidate solvent molecules. A comprehensive set of 215 descriptors was generated for each of the PubChem and generated amines. After correlation and variance filtering, a compact and non-redundant set of descriptors was produced for both chemical spaces, shown in correlation heatmaps (Fig 6, 7). The generated amine space retained 7: MolLogP (lipophilicity), TPSA (topological polar surface area), NumHDonors (H-bond donors), MaxPartialCharge (electronic distribution), Ipc (molecular complexity), Kappa3 (shape index), and BertzCT (structural complexity). The PubChem space retained an identical set except MolWt (molecular weight) replaced BertzCT.

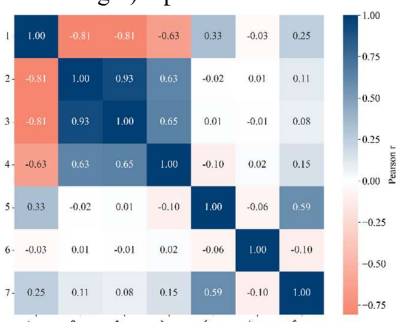


Figure 2. Correlation heatmap for the generated amine dataset, illustrating pairwise Pearson correlations (low [blue] to high [red]). Descriptor Key: 1: MolLogP, 2: TPSA, 3: NumHDonors, 4: MaxPartialCharge, 5: Ipc, 6: Kappa3, 7: BertzCT

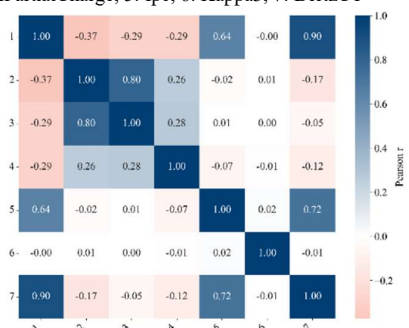


Figure 3. Correlation heatmap for the PubChem amine dataset, illustrating pairwise Pearson correlations, Descriptor Key: 1: MolLogP, 2: TPSA, 3: NumHDonors, 4: MaxPartialCharge, 5: Ipc, 6: Kappa3 7: MolWt

3.2 Training Set Selection

Following robust z-score standardisation, 49 molecules from the generated amine dataset and 766 molecules from the PubChem dataset were excluded as outliers ($Z > 3$ for at least one descriptor). The Elbow method identified optimal configurations of 2 principal components and 10 clusters for the generated amines, and 2 principal components and 12 clusters for the PubChem amines. A further 2 generated amines and zero PubChem amines were removed as PC2 outliers ($PC2 > 10$), and no small clusters (< 10 molecules) existed. An initial D-optimal set of 50 generated amines and 250 PubChem amines was obtained from the Fedorov exchange algorithm, applied independently to each remaining cluster with proportional quotas ($q = 1 \times p$). The overall D-score for the generated was 19.982 and 18.740 for the PubChem amine dataset. Following the gPROMS feasibility test, 22 generated amines and 226 PubChem amines were excluded due to failure to reach VLE convergence, resulting in two processed training sets of 28 feasible generated amines and 24 feasible PubChem amines. The PCA plots show the distribution of candidate molecules in the first two principal components for both design spaces, with D-optimal selected molecules highlighted (Fig. 4). PC1 accounts for 44.3% of the variance in the generated set and 72.3% in the PubChem set, while PC2 accounts for 31.1% and 11.3%, respectively.

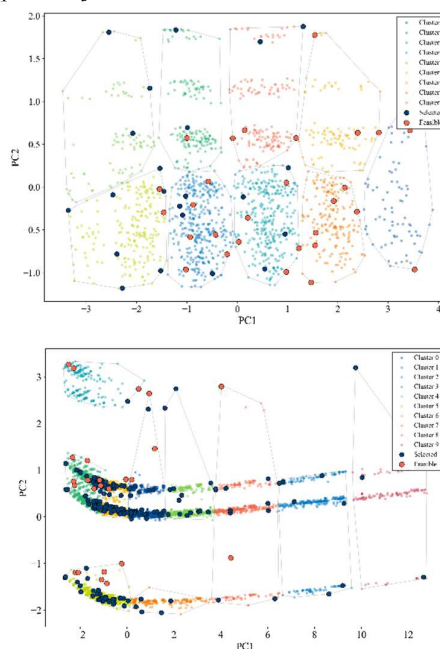


Figure 4. PCA plots for the generated (top) and PubChem (bottom) amine datasets, clustered using k-means ($k = 10$). Voronoi regions outline cluster boundaries, with selected molecules in dark blue and feasible molecules in red. After combining the two feasible datasets, a final feasible training set consisting of 52 solvent molecules was created, with a combined D-

optimality score of 31.161. The distributions of feasible molecules for both datasets show descriptor space coverage (Fig.5). Descriptor coverage of the combined descriptor space was found to be 54.4% for the generated amines and 94.3% for the PubChem amines.

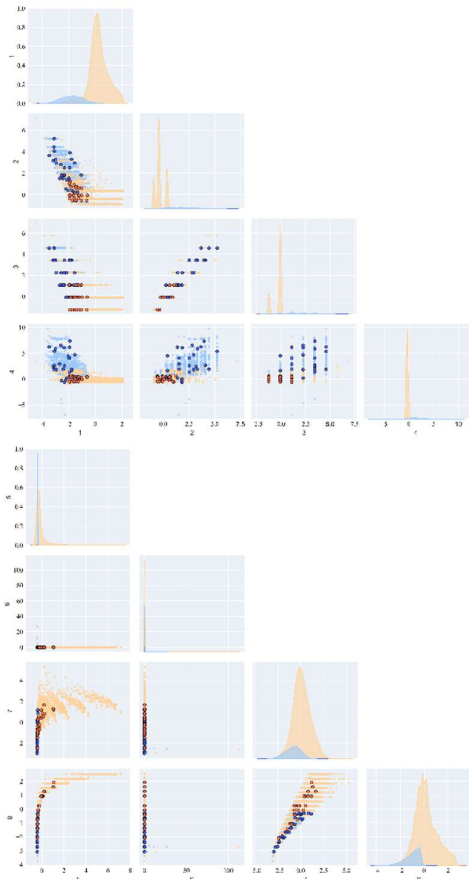


Figure 5. Pairplots showing the descriptor distributions of feasible molecules (black outlined) from the generated (blue) and PubChem (orange) datasets. Descriptor key: 1: MolLogP, 2: TPSA, 3: NumHDonors, 4: MaxPartialCharge, 5: Ipc, 6: Kappa3, 7: BertzCT, 8: MolWt

3.3 ANN Training and Evaluation

Evaluation metrics and parity plots for all outputs for Model 2 (test set) and out-of-sample MEA validation are displayed (Table 3, Fig. 6, 7), with red highlighting outputs not meeting the required performance thresholds (Section 2.5). The predicted outputs include the liquid mole numbers (n_{xi}), liquid compositions (x_i), vapour mole numbers (n_{yi}), vapour compositions (y_i) for H_2O , N_2 , the absorbent (A), and CO_2 , as well as the vapour-to-liquid ratio (β). Overall MAE and RMSE for Models 1,2, MEA validation and training set are shown in Table 4. The test set comprised 10% of datapoints from the 52 training amines across different operating conditions, while MEA represents out-of-sample validation on an unseen amine structure (16,384 datapoints). The number of datapoints per parameter in Model 2 was 5.25

(1,62173 parameters), and 2.95 for Model 1 (72300 parameters).

output	Model 2			MEA Validation		
	MAE	RMSE	R ²	MAE	RMSE	R ²
n_{XH_2O}	0.034	0.133	0.784	0.022	0.063	0.957
n_{XA}	0.002	0.007	0.943	0.003	0.010	0.955
n_{XCO_2}	0.001	0.004	0.984	0.006	0.009	0.807
n_{XN_2}	0.000	0.001	-5050	0.000	0.000	0.981
x_{H_2O}	0.043	0.154	0.788	0.038	0.079	0.900
x_A	0.021	0.067	0.893	0.049	0.142	-0.123
x_{CO_2}	0.028	0.097	0.672	0.024	0.058	0.003
x_{N_2}	0.000	0.004	-579	0.000	0.000	0.474
n_{yH_2O}	0.050	0.153	0.658	0.022	0.063	0.938
n_{yA}	0.000	0.001	0.978	0.003	0.010	0.249
n_{yCO_2}	0.001	0.002	0.996	0.006	0.009	0.974
n_{yN_2}	0.000	0.001	1.000	0.000	0.000	1.000
y_{H_2O}	0.047	0.139	0.746	0.215	0.396	-0.592
y_A	0.000	0.003	0.882	0.005	0.010	0.198
y_{CO_2}	0.007	0.036	0.976	0.127	0.272	0.396
y_{N_2}	0.041	0.130	0.833	0.005	0.012	0.998
β	0.028	0.096	0.889	0.026	0.064	0.964

Table 3: Evaluation metrics (MAE, RMSE, R²) for Model 2 and MEA Validation, with red highlighting metrics that did not meet the defined thresholds.

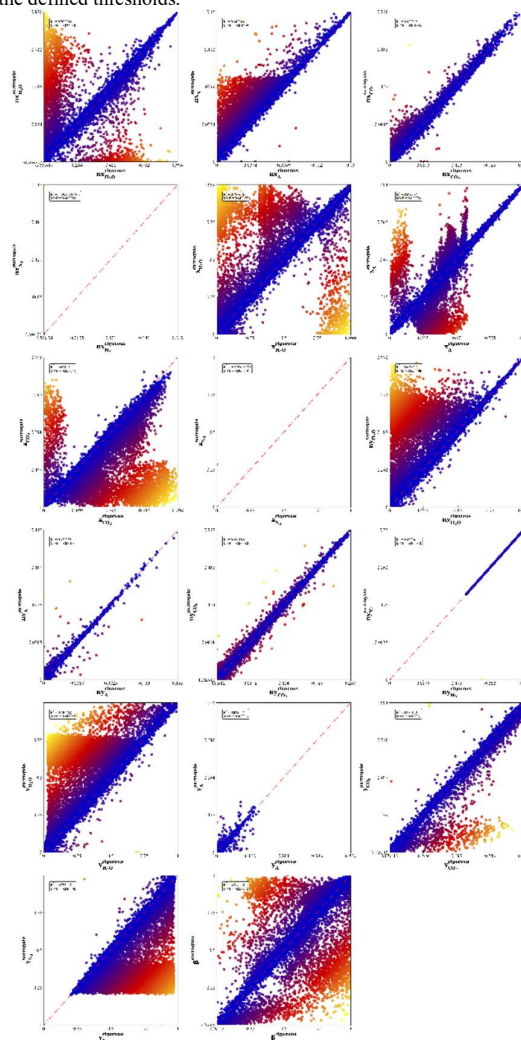


Figure 6: Parity plots for Model 2, showing predicted vs. true values for the 17 VLE outputs. Point colour reflects absolute error, ranging from blue (low) to yellow (high).

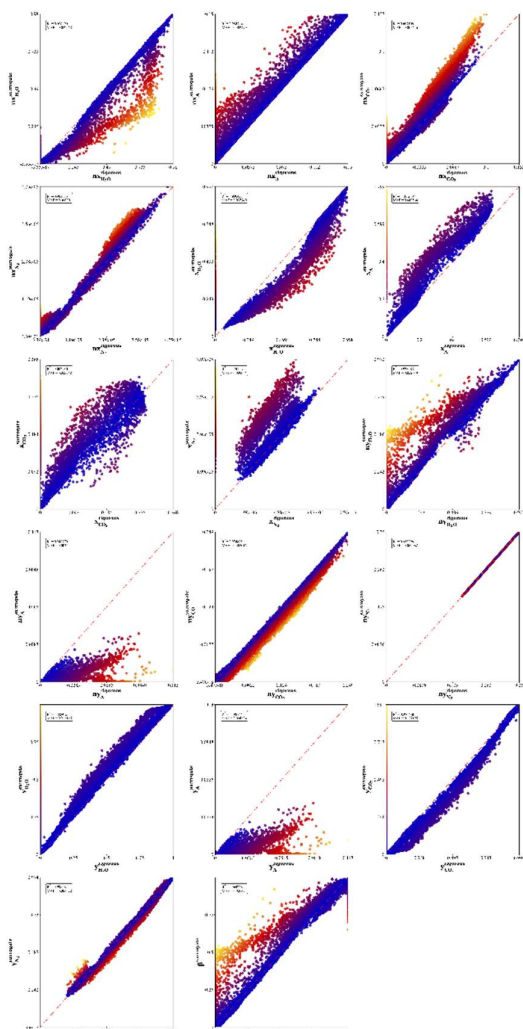


Figure 7: Parity plots for MEA validation, showing predicted vs. true values for the 17 VLE outputs. Point colour reflects absolute error, ranging from blue (low) to yellow (high).

	MAE	RMSE
Model 1	0.0190	0.061
Model 2	0.0179	0.060
MEA Validation	0.0324	0.070
Training Set	0.0159	0.055

Table 3: Overall evaluation metrics (MAE, RMSE) for Models 1, 2, MEA Validation and training set.

4. Discussion

4.1 Design Space Strategy

Two complementary design spaces were used to ensure a computationally feasible and chemically diverse final optimal set of solvent molecules. The first generated space was subject to prescriptive structural and property constraints, guaranteeing the MILP problem was computationally tractable and likely to generate chemically valid GC vectors. However, the presence of these constraints limited the diversity of the design space, particularly excluding heavier polyamines and highly functionalised structures that could potentially exhibit optimal CO₂ performance. To address this, a second selection space was constructed using real

and experimentally verified chemical amine structures obtained from PubChem, which was not subject to the same prescriptive constraints as the generated space. This approach yielded a more diverse and varied selection of GC vectors, including polyamines and amino alcohols (larger descriptor coverages as discussed in Section 4.3 (Fig.5)), despite lower VLE feasibility rates. This approach provided training data with dense coverage of industrially relevant solvents from the generated space, while incorporating broader structural diversity from the PubChem space, exposing the model to less typical solvents with potentially high CO₂ capture performance. The effectiveness of this combined solvent space strategy has been evaluated quantitatively using design space coverage in Section 4.3 and ultimately through model performance.

4.2 Molecular Descriptor Calculation and Screening

The final training dataset required a representative, diverse sample of the molecular space. A compact, meaningful set of descriptors that encapsulate the structural and electronic properties, and therefore, factors determining amine performance during CO₂ capture (i.e. high absorption rates and capacity, good CO₂ loading), was obtained before sampling. The descriptor set was filtered to reduce dimensionality while minimising information loss by eliminating features with standard deviations $< 1 \times 10^{-3}$ and one descriptor from each pair with a correlation > 0.97 ; an established approach to improve sampling diversity and efficiency.¹² Both design spaces retained exactly 7 descriptors after screening (within our target range of 6–10), collectively representing solubility characteristics (MolLogP, TPSA), H-bonding capacity (NumHDonors), molecular topology (Ipc, Kappa3), and electronic properties (MaxPartialCharge). BertzCT was kept for the generated space, whereas MolWt was kept for PubChem, as MILP constraints made topological complexity more informative than size.

For the generated amine dataset, the correlation heatmap (Fig. 2) show mostly low correlations between descriptors ($|r| < 0.4$), confirming diversity. However, due to the relationship between hydrophobic chain length and polarity, a strong negative correlation between MolLogP and TPSA ($r \approx 0.8$) was observed. NumHDonors shows strong correlation with TPSA ($r \approx 0.93$) and negative correlation with MolLogP ($r \approx -0.81$), indicating that H-bond donors contribute to polar surface area and oppose lipophilicity. The three descriptors NumHDonors, TPSA, and MolLogP together define the main chemical variation axes, ranging from highly polar and hydrophilic primary amines to less polar and more lipophilic tertiary amines. This shows the data effectively covers the CO₂ capture trade-off

between high reactivity/absorption (primary amines) and low regeneration energy requirements (tertiary amines).

The PubChem amine dataset shows weaker overall correlations (Fig. 3), illustrating that real compounds have more diverse structures than the generated compounds. The NumHDonors-TPSA correlation ($r \approx 0.80$) remains high yet reduced compared to the generated space (0.93). The NumHDonors-MolLogP correlation shows a significant decrease in strength, from -0.81 to -0.29, indicating that PubChem amine lipophilicity depends on factors beyond H-donor count, including chain length and branching patterns. Overall, the heatmap shows weaker, more variable relationships than the generated set, reflecting the unfiltered diversity of real-world chemical libraries. Additional observations on descriptor distributions are provided in the SI (Fig SI.1.2) and confirm sufficient diversity for D-optimal sampling (D-score 31.16). While descriptor diversity is necessary for effective training data sampling, it does not guarantee model success as discussed in the following sections.

4.3 Sampling section

The generated amine dataset, constrained by molecular weight (≤ 12 carbons), GCs, and flash point requirements occupied a concentrated and evenly distributed region of descriptor space centred around typical desirable CO₂ capture solvent properties (Fig. 4, 5). These constraints resulted in a higher proportion of the generated amines passing the feasibility VLE convergence tests, (56% vs. 10% for PubChem), but also limited descriptor coverage to 54.4% of the combined range, underrepresenting higher molecular weights, elongated shapes ($K_{\text{p}3} < 5$), and multi-amine structures (NumHDonors > 3).

Conversely, the largely unconstrained PubChem dataset spanned a broader descriptor space with greater variance in PC1 (72.3% versus 44.3% for generated set) achieving 94.3% descriptor coverage (Fig. 4). Despite lower feasibility rates, the PubChem dataset proved valuable for final model training due to increased structural diversity. The D-optimal scores increased from 19.98 (generated) and 18.74 (PubChem) individually to 31.161 for the combined dataset, indicating improved coverage of the design space. Including an approximately equal number of datapoints from each set meant the training data captured a dense sample of industrially relevant solvent properties, along with extended coverage to edge cases absent from the generated space. Initial sampling targeted 50 generated and 250 PubChem amines to yield ~ 25 feasible molecules from each. The target of 50 was chosen due to time constraints/computational limitations of the project. This D-optimal, descriptor-based sampling method ensured exposure to diverse molecular structures

and has been shown to effectively maximise training data information and improve surrogate model performance.¹³ However, the ultimate success of the model depends not only on sampling diversity and size, but whether GC representations can adequately capture VLE behaviour, which is addressed in the following section.

4.4 Surrogate Model Training and Evaluation

This work aimed to learn how molecular structures influence VLE behaviour from 52 diverse amines enabling rapid solvent screening for CO₂ capture. While Baakes *et al.* demonstrated that ANNs can produce highly accurate VLE predictions for a single solvent system (MEA), our multi-solvent CAMPD framework exhibited significantly larger errors. Unlike Baakes' approach, which learned from fixed MEA chemistry across varying conditions, our model's objective was to learn how the structure affects VLE properties over various amine chemistries.

From the evaluation of the test set, we can see this objective was partially achieved with four outputs (of 17) fully meeting all thresholds, showing parity plots with high concentrations of points along the diagonal line: CO₂ liquid mole numbers (n_{CO_2}), amine, CO₂ and N₂ vapour mole numbers (y_{A} , y_{CO_2} , y_{N_2}). Vapour CO₂ predictions and all amine predictions were particularly accurate ($R^2 > 0.97$), representing a core carbon capture metric. Notably, y_{N_2} shows a perfect $R^2 = 1.000$, which likely reflects the simple behaviour of inert and highly volatile N₂, rather than an indication of sophisticated learning.

However, systematic failures occurred in H₂O predictions (R^2 0.66-0.78 across phases) and liquid compositions (x_{CO_2} R^2 0.67, x_{A} R^2 0.89), with β ratio also failing to meet thresholds. Liquid N₂ predictions in Model 2 show catastrophic R^2 values (n_{N_2} : -5000, x_{N_2} : -500) with parity plots clustered at near-zero; these results occur as the true values are very low across all conditions, with a near-zero variance. Absolute errors remain small, indicating accurate predictions despite R^2 instability. Model 1 avoids this issue, suggesting Model 2's larger architecture overfits to numerical noise in trace components. Vapour amine (y_{A}) exhibits a similar trace component issue with failing R^2 thresholds (0.88-0.91) despite excellent absolute errors (MAE ≈ 0.000), as vapour amine concentrations remain low due to low volatility.

Three preliminary models with varying architectures were tested on the smaller dataset to identify suitable hyperparameters (details in SI). The most effective configuration (100 hidden units, 8 layers, 300 epochs) was selected as Model 1. For Model 2, the architecture was scaled to 150 hidden units with 250 epochs to accommodate the larger dataset. Model 2 showed a marginal improvement in overall performance (MAE: 0.0179; RMSE: 0.06, compared with 0.0190 and 0.061 for Model

1). However, no substantial gains were observed despite 4x more training data, and the same outputs failed in both models. Both models operated below the recommended ~10 datapoints per hyperparameter (Model 2: 2.86, Model 1: 5.19), typically indicating an overfitting risk. Training set evaluation (MAE: 0.0159, RMSE: 0.0554) showed marginal improvements in performance over the test set (MAE: 0.0179, RMSE: 0.0600), with the same outputs failing in both sets (H_2O predictions (R^2 0.66-0.78), liquid compositions ($x\text{CO}_2$ R^2 0.67), and β ratio). This minimal difference suggests that overfitting or data scarcity for each amine was not the primary limitation. Additionally, while time constraints prevented a full hyperparameter optimisation, tuning may have yielded incremental gains, but it is unlikely to resolve these fundamental challenges.

Instead, the model's poor performance was likely due to insufficient sample size and/or inadequate feature representation. The GCs assume additive contributions, which may not capture sufficient structural information to predict complexities in H_2O -amine hydrogen bonding, where the same functional group behaves differently depending on steric accessibility and neighbouring groups, leading to confusing results. With only 52 training amines spanning a 19-dimensional input space, the model likely lacked enough examples to learn these complex structure-VLE relationships. This highlights a key difference from SAFT- γ Mie, which succeeds with GC representations by combining them with physics-based equations for hydrogen bonding and thermodynamic relationships. In contrast, our data-driven ANN must learn these complex interactions empirically. Reviews of machine learning for CO_2 capture materials (including both solid absorbents and liquid solvents) have identified that charge-based, orbital-based, and thermodynamic descriptors are required to capture CO_2 interactions effectively, supporting that GC molecular representations may be insufficient to predict complex VLE behaviour.¹⁴

While the test set evaluation assessed model accuracy across operating conditions for the 52 training amines, MEA out-of-sample validation tested model generalisation to an unseen molecular structure, the industrial benchmark primary monoamine. Overall, MAE increased 81%; however, it did show selective improvements from the test set: water mole numbers substantially improved ($n\text{xH}_2\text{O}$: R^2 0.78 to 0.96, $n\text{yH}_2\text{O}$: R^2 0.66 to 0.94), β ratio predictions strengthened (R^2 0.89 to 0.96), and the trace N_2 artefact was resolved ($n\text{xN}_2$: R^2 -5050 to 0.98). These gains likely reflect MEA's simpler primary monoamine structure, making VLE behaviour easier to predict. Critically, however, compositional predictions catastrophically failed: liquid amine composition

collapsed to R^2 -0.12, liquid CO_2 to R^2 0.00, and vapour compositions degraded significantly ($y\text{H}_2\text{O}$: R^2 -0.59, $y\text{CO}_2$: R^2 0.40). Parity plots confirm this pattern, showing tight diagonal clustering for mole numbers but scatter for compositions (Fig. 7).

The consistently worse predictions for all compositions, relative to their corresponding absolute quantities, are suspected to be from the surrogate's architecture. The absolute quantities are linearly reconstructed, and the compositions are then obtained by normalising these mole numbers; this introduces nonlinear sensitivity to relative errors among species. The vapour fraction is most affected by this reconstruction step as it is dependent on the total liquid-phase molar amount, resulting in error accumulation and greater inaccuracies. Additionally, the RMSE loss treats all 17 reconstructed VLE outputs equally, despite their differing complexity. Since compositions and β arise from nonlinear transformations, whereas mole numbers are linear, this uniform weighting leads to unequal accuracy across variables. Σn_x is dominated by $n_{\text{xH}_2\text{O}}$, this bias also propagates into β , contributing to the V-L ratio failing the accuracy thresholds. These failures along with the V-L ratio not passing the defined metrics, limit the current model's utility. Despite accuracy limitations, the results demonstrate that multi-solvent CAMPD is feasible but currently limited by molecular representation, data scarcity, and model architecture.

5. Conclusions

The objective of the project was to extend the single-solvent surrogate model approach for the SAFT- γ Mie EoS (Baakes *et al.*³) to a multi-solvent CAMPD framework that predicts VLE data from GC inputs, using 52 diverse amines selected with k-means clustering and D-optimal sampling for training, testing and validation. Two multi-solvent frameworks (4096 and 16,384 datapoints per amine) were developed, providing inherent computational benefits over the rigorous SAFT- γ Mie EoS calculation. Four of 17 output metrics were predicted successfully, including the core metric (vapour CO_2 $R^2 > 0.97$). Despite this success, H_2O predictions failed across both phases, and liquid-phase and β predictions exhibited insufficient accuracy for process design. MEA out-of-sample validation reinforced these limitations: while water mole number predictions improved for this simpler primary monoamine structure, compositional distributions failed catastrophically (liquid amine R^2 -0.12, liquid CO_2 R^2 0.00), demonstrating the model cannot generalize compositional behaviour to unseen chemistries. We identified the primary limitations for the accuracy of the model as insufficient sample size, with 52 amines not enough to map a 19-dimensional input space to VLE outputs, and molecular representation issues with the GC

method, with results suggesting GCs may not provide sufficient structural information to predict liquid-phase VLE properties effectively. Time constraints also meant a full hyperparameter optimisation was not possible, which could also limit current accuracy. Our results demonstrate that a multi-solvent CAMPD is feasible for pre-screening of amine solvents, with the rigorous SAFT validation still required for final candidates. Future work should focus on increasing molecule sample size (increase from 52 to 200+ amines), full hyperparameter optimisation, and the potential

integration of molecular descriptors that are likely to be more informative for the liquid-phase and H₂O predictions (hydration descriptors). Adaptive sampling strategies, i.e. active learning, where the model guides the selection of additional training molecules to sample regions of high prediction uncertainty, should also be looked at. Overall, our work represents the first multi-solvent surrogate framework for amine-based carbon capture, identifying both feasibility and key challenges for future development.

References

- ¹ IEA (2025) 'CO₂ Emissions in 2024', *Global Energy Review 2024*. Available at: <https://www.iea.org/reports/global-energy-review-2025/co2-emissions>.
- ² Wilcox, J. (2012). *Carbon Capture*. [online] Google Books. Available at: <https://doi.org/10.1007/978-1-4614-2215-0>
- ³ Baakes, F., Chaparro, G., Bernet, T., Jackson, G., Galindo, A. & Adjiman, C.S. (2025) *Utilizing ML Surrogates in CAPD: Case Study of an Amine-based Carbon-Capture Process*. Systems and Control Transactions. Available at: <https://psecommunity.org/LAPSE%3A2025.0327v1>
- ⁴ Lu, G., Wang, Z., Bhatti, U.H. & Fan, X. (2023) 'Recent progress in carbon dioxide capture technologies: A review', *Clean Energy Science and Technology*, 1(1), p. 32. Available at: <https://doi.org/10.18686/cest.v1i1.32>
- ⁵ Bui, M., Adjiman, C.S., Bardow, A., Anthony, E.J., Boston, A., Brown, S. & Fennell, P.S. (2018) 'Carbon capture and storage (CCS): the way forward', *Energy & Environmental Science*, vol. 11, pp. 1062–1176. Available at: <https://eprints.whiterose.ac.uk/id/eprint/128986/1/c7ee02342a.pdf>
- ⁶ Lee, Y.S., Jackson, G., Galindo, A. & Adjiman, C.S. (2023) 'A predictive model for the techno-economic assessment of CO₂ chemisorption processes applicable to a large number of amine solvents', *Chemical Engineering Research and Design*, 200, pp. 615–636. Available at: <https://doi.org/10.1016/j.cherd.2023.11.007>
- ⁷ Lee, Y.S., Galindo, A., Jackson, G. & Adjiman, C.S. (2023) Enabling the direct solution of challenging computer-aided molecular and process design problems: Chemical absorption of carbon dioxide. *Computers & Chemical Engineering*, 174, 108204. Available at: <https://doi.org/10.1016/j.compchemeng.2023.108204>
- ⁸ McBride, K. & Sundmacher, K. (2019) Overview of surrogate modelling in chemical process engineering. *Chemie Ingenieur Technik*, 91(3), pp. 228–239. Available at: <https://doi.org/10.1002/cite.201800091>
- ⁹ Dobbelaere, M.R., Plehiers, P.P., Van de Vijver, R., Stevens, C.V. & Van Geem, K.M. (2021) 'Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats', *Engineering*, 7, pp. 1201–1211. Available at: <https://doi.org/10.1016/j.eng.2021.03.019>
- ¹⁰ Amar, Y., Schweidtmann, A.M., Deutsch, P., Cao, L. and Lapkin, A. (2019) 'Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis', *Chemical Science*, 10(27), pp. 6697–6706. Available at: <https://doi.org/10.1039/C9SC01844A>
- ¹¹ Zhang, X., Erdős, D.L., Ahunbay, M.G., Chen, B., Doherty, M.F. and Bardow, A. (2021) 'Integrated ionic liquid and rate-based absorption process design for gas separation: Global optimization using hybrid models', *AIChE Journal*, 67(5), e17340. Available at: <https://doi.org/10.1002/aic.17340>
- ¹² Gramatica, P. (2020) 'Principles of QSAR modeling: comments and suggestions from personal experience', *International Journal of Quantitative Structure-Property Relationships*, 5(3), pp. 61–90. Available at: <https://doi.org/10.4018/IJQSPR.20200701.oa1>
- ¹³ Gui, L., Armstrong, A., Galindo, A., Sayyed, F.B., Kolis, S.P. and Adjiman, C.S. (2024). On the design of optimal computer experiments to model solvent effects on reaction kinetics. *Molecular Systems Design & Engineering*, 9(12), pp.1254–1274. doi:<https://doi.org/10.1039/d4me00074a>.
- ¹⁴ Orhan, I.B., Zhao, Y., Babarao, R., Thornton, A.W. and Le, T.C. (2025). Machine Learning Descriptors for CO₂ Capture Materials. *Molecules*, [online] 30(3), p.650. doi:<https://doi.org/10.3390/molecules30030650>.

CFD of Blood Flow in Marfan Patients: A Haemodynamic Comparison between Medical Management and Aortic Root Replacement Surgery

Mason Liu, Shaozhi Huang

Department of Chemical Engineering, Imperial College London, U.K.

Abstract — Marfan patients are often left with the only option of aortic root replacement to prevent potential dissection when the degree of ascending aorta dilation exceeds the surgical threshold. However, successful replacement sometimes leads to downstream implications such as distal aortic aneurysms and dissection in the proximal descending aorta. The aim of this study was to assess the risks involved with surgical replacement by comparing blood flow and haemodynamic metrics with conservative treatment, using Phase-Contrast Magnetic Resonance Imaging (PC-MRI) and Computational Fluid Dynamics (CFD). MR images from two patients were acquired: one who underwent medical treatment and one who underwent aortic root and ascending aorta replacement, and patient-specific flow conditions were implemented into CFD simulations. The results revealed that flow velocity and Time-Averaged Wall Shear Stress (TAWSS) had drastically increased after surgical procedures, accompanied by a 350% increase in viscous energy loss. Stronger vortical flow overall and especially in the distal descending aorta was also discovered. Analysis suggests that abnormal haemodynamics after replacement surgeries contribute to increased risk of aneurysm growth and type-B dissection at an early stage, implying the potential to advance downstream risk stratification by improving preliminary modelling before surgery.

1 Introduction

Marfan syndrome (MFS) is a connective-tissue disorder caused by pathogenic variants in the *FBN1* gene encoding fibrillin-1, which leads to abnormal microfibril architecture [1]. Cardiovascular involvement occurs in 90% of patients and frequently manifests as aortic root dilation, thoracic aortic aneurysms (TAAs), and, in the worst case, aortic dissection [2][3]. Clinically, risk assessment of MFS mainly relies on morphological features and population-derived risk scores. Surgical replacement of the aortic root and ascending aorta is typically recommended when the sinus diameter exceeds 50 mm [4]. However, despite successful root replacements, many patients still develop distal aortic aneurysms or dissections at a later stage [5], suggesting that current morphological criteria do not fully capture the haemodynamic mechanisms driving disease progression, and that more accurate stratification is required before surgery.

The analysis of blood flow in aortas has been facilitated by advances in imaging techniques and computational fluid dynamics (CFD). While anatomical MRI captures aorta geometry with high resolution, time-resolved phased-contrast MRI (PC-MRI) quantifies blood velocity through specified imaging planes [6]. Several studies have incorporated data from PC-MRI into patient-specific CFD simulations [7] to investigate features such as wall shear stress (WSS), vortices, and helices in patients with TAAs [8][9], dissections [10][11], and replaced aortic roots [5][12]. The visualisation of haemodynamics may highlight alterations before and after surgery, informing potential risks of aneurysm growth and onset of dissection [13].

The possibility of haemodynamic factors serving as predictive markers of aneurysm growth and root dilation have been widely explored. WSS is a three-dimensional vector that represents the tangential forces exerted by blood

flow on vessel walls as a result of complex flow structures. In patients with MFS, abnormal spatial distribution and temporal variations of WSS along the ascending aorta and arch have been associated with regions of eccentric jets and secondary flow [14]. Areas that later undergo accelerated growth often coincide with segments of either elevated WSS, which is thought to promote adverse vascular remodelling and matrix degradation [15], or steep spatial WSS gradients and elevated oscillatory shear index (OSI), which may contribute to endothelial dysfunction. Helical and vortical flows are other occurrences that distinguish Marfan from healthy aortas. Increased vortical flow has been reported in the descending aorta and is associated with dilation in the sinus [13][16]. This is thought to cause disease progression by altering shear stress forces at the vessel wall and creating areas of increased risk [9]. These irregular blood flow patterns are often shown at an earlier stage of disease progression, thus personalised in-vitro modelling may help identify these risks.

Beyond these commonly studied haemodynamic metrics, the influence of energy loss in the aorta on the progression of Marfan syndrome has received less focus, largely due to limited understanding of the underlying mechanisms. However, it has been suggested that when mechanical energy in blood flow transforms into thermal energy, the morphology and physiological function of endothelial cells on the vessel wall can be impaired, possibly promoting vascular diseases [17]. In the aorta, the majority of energy loss occurs through viscous dissipation, while turbulence dissipation and mutual work between the aortic wall and blood flow contribute to a smaller proportion [18]. Thus, it is important to quantify abnormal patterns or elevations in dissipative losses, which may present an additional marker of adverse remodelling and further dilation.

The primary aim of this study is to determine how

different treatment strategies in MFS, notably conservative management and aortic root replacement, influence aortic haemodynamics and potential markers of distal disease risk. Using patient-specific CFD with multi-modality imaging data, flow velocity patterns, WSS distributions and helicity structures will be quantified, and energy losses through dissipation will also be evaluated. Comparative analysis of these haemodynamic parameters between treatment strategies and time points will help identify whether surgical intervention mitigates or redistributes adverse conditions along the aorta, and whether energy loss provides additional insight into the risk of distal aortic remodelling.

2 Methodology

2.1 Data Acquisition

Phase-contrast Magnetic Resonance Imaging (PC-MRI) data were acquired from two Marfan patients who received different treatment methods. Patient 1 received conservative medical treatment and had been routinely monitored. Data from two time points (2022 and 2024) were selected to track disease progression. Patient 2 underwent aortic root and ascending aorta replacement in 2017 after the dilated ascending aorta diameter exceeded the surgical threshold for less invasive measures such as Personalised External Root Support (PEARS). Notably, the surgical graft used for this patient was kinked. Both pre-surgery (2012) and post-surgery (2017) scans were available for this patient. The inlet surface was created by segmentation at the sinotubular junction (STJ), where the PC-MRI flow mapping plane lies, in order to accurately implement patient velocity components. Names of parts and sections in the aorta referred to in this study are outlined in Fig. 1.

2.2 Mesh Generation

Computational meshes were generated using ANSYS ICEM CFD (v.2024R2). All meshes were unstructured and consisted of tetrahedral cores, with ten prism layers reducing in size towards the aortic wall to ensure adequate boundary resolution [7]. Local areas of higher mesh density were created around sharp bends near the supra-aortic branches and aneurysms in the proximal descending aorta. The density regions were kept constant for each patient, ensuring accurate comparison between meshes. The Octree method was first employed, followed by Delaunay and the addition of prism layers. Mesh smoothing was performed after completion of each step, and all final meshes had a mesh quality over 0.3.

2.3 Sensitivity Analysis

Sensitivity analysis was conducted to ensure the solution is mesh-independent. Transient simulations with a generic waveform were performed under a flat velocity profile normal to the inlet plane. The mean and maximum velocity magnitude and time-averaged wall shear stress (TAWSS) were calculated at 4 representative cross-sections along the aorta where the highest gradients in these variables were observed. The mesh was iteratively refined until variations in these variables between successive meshes were less than 3.5%. The grid convergence index (GCI) was computed to quantify residual discretisation error, and

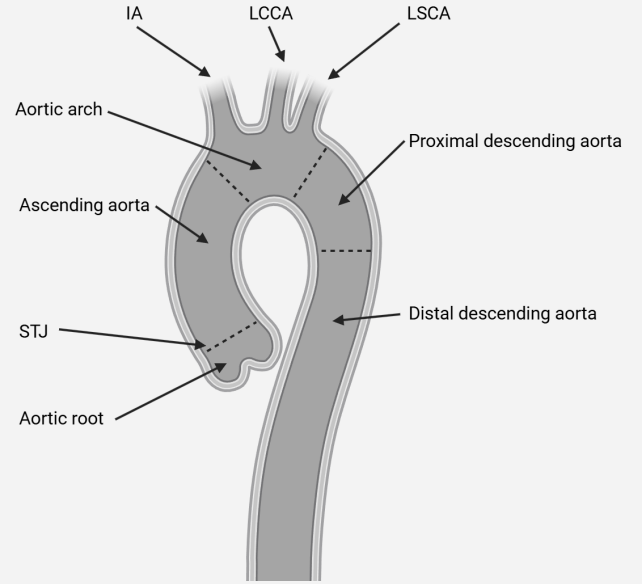


Figure 1: Terminology of parts and sections in the aorta. STJ = sinotubular junction, IA = innominate artery, LCCA = left common carotid artery, LSCA = left subclavian artery.

the final selected mesh exhibited a GCI below 5.5% [7]. The detailed calculations can be found in Supplementary Information A.

2.4 Fluid Model

The aorta walls were defined as impermeable and rigid with no-slip conditions imposed. Blood was modelled as a non-Newtonian, incompressible fluid with a density of $1060 \text{ [kg m}^{-3}\text{]}$ and viscosity governed by the Carreau-Yasuda model [19]:

$$\mu(\dot{\gamma}) = \mu_{\infty} + (\mu_0 - \mu_{\infty})[1 + (\lambda\dot{\gamma})^a]^{\frac{(n-1)}{a}} \quad (1)$$

where μ_{∞} and μ_0 are the infinite and zero shear viscosity with values of 0.0035 [Pa s] and 0.16 [Pa s] , respectively, $\dot{\gamma}$ is the shear rate, λ is the time constant at 8.2 [s] , a is the transition parameter with a value of 0.64 , and n is the power law exponent with a value of 0.2128 .

Time-dependent pressure outlet boundary conditions were based on the Three-Element Windkessel Model (3-EWM), with its parameters tuned using patient blood pressure and geometry data. It was assumed that outflow from the descending aorta accounted for 70% of total blood flow, while the remaining 30% was partitioned between the three supra-aortic branch outlets (IA, LCCA, LSCA) based on their cross-sectional area. Brachial blood pressures were measured as $111/72$ (2022) and $114/71$ (2024) [mmHg] for Patient 1, and $121/61$ (pre) and $133/74$ (post) [mmHg] for Patient 2, and the mean arterial blood pressure, p_m , was calculated from:

$$p_m = p_{dia} + \frac{1}{3}(p_{sys} - p_{dia}) \quad (2)$$

where p_{dia} and p_{sys} denotes diastolic and systolic brachial pressures, respectively. Table 1 summarises the Windkessel parameters which were calculated as described in Pirola et al. [20]. Details of the calculations can be found in Supplementary Information C.

Patient	Vessel	R ₁ [10 ⁸ Pa s m ⁻³]	R ₂ [10 ⁸ Pa s m ⁻³]	C [10 ⁻⁹ m ³ Pa ⁻¹]
P1-2022	DA	0.146	2.955	5.771
	IA	0.630	11.233	1.509
	LCCA	2.746	39.895	0.420
	LSCA	2.035	30.824	0.545
P1-2024	DA	0.160	2.982	5.696
	IA	0.695	9.867	1.695
	LCCA	5.249	5.601	0.292
	LSCA	3.160	6.242	0.454
P2-PRE	DA	0.222	2.303	7.088
	IA	1.450	14.582	1.117
	LCCA	1.080	11.334	1.442
	LSCA	3.834	33.509	0.479
P2-POST	DA	0.218	1.450	10.735
	IA	0.682	5.998	2.680
	LCCA	5.907	37.749	0.410
	LSCA	1.318	10.530	1.511

Table 1: 3-EWM Parameters for Aortic Outlets. P1 and P2 refer to Patients 1 and 2, respectively. R1 = proximal resistance, R2 = distal resistance, C = compliance, IA = innominate Artery, DA = descending aorta, LCCA = left common carotid artery, LSCA = left subclavian artery.

2.5 Simulation Setup

All simulations were performed using ANSYS Fluent (v.2024R2) with a pressure-based transient solver. The Windkessel and Carreau-Yasuda models were prescribed through user-defined functions, detailed in Supplementary Information B. The inlet velocity profile was implemented with point-wise velocity component data at the inlet plane. The SIMPLE pressure-velocity coupling scheme was employed, and second-order pressure and time discretisation schemes were applied to ensure accurate transient simulation. A normalised residual requirement of 10^{-4} was imposed on velocity and pressure. A time step of 0.001 [s] was used, and each time step was limited to a maximum of 50 iterations, which proved sufficient for the solution to converge during trial runs.

Based on calculations of the peak Reynolds number (Re_{peak}) and the Womersley parameter (α) in the ascending aorta (Table 2), which were compared against the stability criteria outlined by Kousera et al. [21] ($Re_{crit} = 150 \alpha$), all cases were overall laminar except for Patient 2 post-surgery. To maintain consistency and to allow for direct comparison between cases without model bias, it was deemed appropriate to compute with a laminar model for all cases. All simulations were run for 3 cardiac cycles to ensure that the solution was periodic, and the last cycle was used for post-processing.

Case	α	Re_{peak}	Re_{crit}	Turbulence
P1-2022	30.79	3964	4618	No
P1-2024	33.74	2117	3828	No
P2-PRE	30.82	3823	4622	No
P2-POST	27.70	5798	4155	Yes

Table 2: Characterisation of flow regimes. Re_{crit} is the critical Reynolds number above which the flow becomes transitional.

2.6 Post-processing

Simulation data was exported in Enight Gold binary format for post-processing in ANSYS Enight. The primary haemodynamic parameters of interest are helicity density (H_k), Wall Shear Stress (WSS), Oscillatory Shear Index (OSI) and energy loss via viscous dissipation (E_V). Helicity density measures the correlation between velocity and its curl, vorticity, and describes the propensity of flow to form vortices or coherent structures [22]:

$$H_k = \mathbf{u} \cdot \boldsymbol{\omega} \quad (3)$$

For better comparison of the degree of vortical flow forming in a fluid domain, helicity is often normalised with respect to the magnitude of velocity and vorticity:

$$H_{k,norm} = \frac{\mathbf{u} \cdot \boldsymbol{\omega}}{|\mathbf{u}||\boldsymbol{\omega}|} \quad (4)$$

In pulsatile flow, shear stress on vessel walls is often represented by the Time-Averaged Wall Shear Stress (TAWSS), computed by integrating the wall shear stress magnitude over the full cardiac cycle (T_c):

$$TAWSS = \frac{1}{T_c} \int_0^{T_c} |\boldsymbol{\tau}_W| dt \quad (5)$$

To quantify the directional change in WSS, the Oscillatory Shear Index (OSI) can be calculated. It is a comparison between the magnitude of the time-averaged shear stress vector and TAWSS:

$$OSI = \frac{1}{2} \left(1 - \frac{|\int_0^{T_c} \boldsymbol{\tau}_W dt|}{\int_0^{T_c} |\boldsymbol{\tau}_W| dt} \right) \quad (6)$$

To understand the relationship between irregular aorta features and extra power requirements for the heart, energy loss is computed. The primary energy loss is through viscous dissipation (ϕ), which accounts for the conversion of mechanical energy into heat:

$$\phi = \tau_{ij} S_{ij} = 2\mu S_{ij} S_{ij} = \frac{\mu}{2} \sum_{i,j} \left(\frac{\delta u_i}{\delta x_j} + \frac{\delta u_j}{\delta x_i} \right)^2 \quad (7)$$

Where τ_{ij} and S_{ij} are the viscous stress tensor and rate of strain tensor, respectively. Integrating the viscous dissipation function over the entire aortic volume yields the viscous energy loss rate, P_V , in Watts [W]:

$$P_V = \int \frac{\mu}{2} \sum_{i,j} \left(\frac{\delta u_i}{\delta x_j} + \frac{\delta u_j}{\delta x_i} \right)^2 dV \quad (8)$$

The net viscous energy loss (E_V , in J) is found by integrating P_V over the full cardiac cycle.

3 Results

3.1 Patient Features

The reconstructed and segmented geometries used in this study are shown in Fig. 2. For Patient 1 under conservative treatment, the diameter of the ascending aorta at the widest point has grown from 23.0 [mm] in 2022 to 31.3 [mm] in 2024. There was also a slight bending of the descending

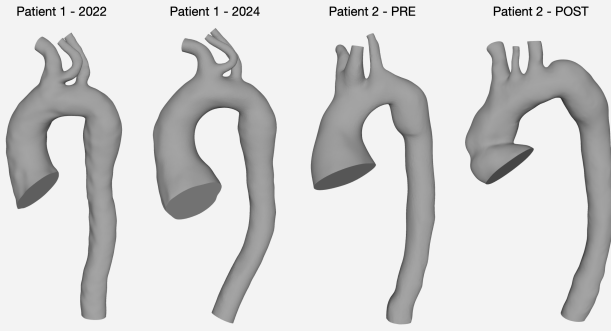


Figure 2: Reconstructed geometries from MR images after segmentation.

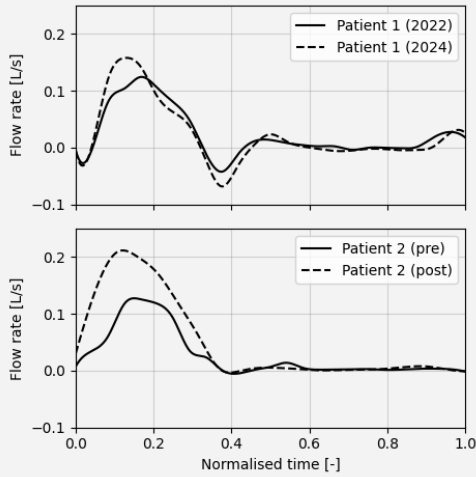


Figure 3: Flow waveforms extracted from PC-MRI data. The time axis has been normalised for better comparison.

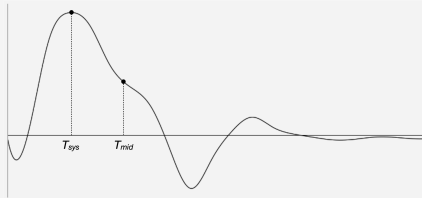


Figure 4: A representation of the position of T_{sys} and T_{mid} during the cardiac cycle.

aorta in 2024. Patient 2 had a visually larger dilation pre-surgery at 28.7 [mm], which was adjusted to 18.8 [mm] through replacement of the aortic root and ascending aorta. The kinked graft could be clearly seen in ascending aorta, with a diameter of 15.4 [mm] at the narrowest point. Both patients showed development of aneurysms downstream of the aortic arch. While the aneurysm in Patient 1 did not grow, the aneurysm in Patient 2 had grown upwards post-surgery.

The flow waveforms extracted from PC-MRI data are shown in Fig. 3. Similar waveforms were seen for Patient 1 across the two years, with only modest differences in peak magnitude and the maximum depth of the early reverse-flow region. In contrast, the pre- and post-surgery waveforms for Patient 2 demonstrated strong alterations, with the post-operative curve exhibiting a significantly lower systolic peak. Despite these differences in peak flow, all waveforms

shared consistency in the timing of key phases, allowing for direct comparison between treatment methods.

3.2 Flow Patterns

The instantaneous velocity streamlines at peak systole (T_{sys}) and mid-systolic deceleration (T_{mid}) were visualised and shown in Fig. 5. A representation of the position of these time points is shown in Fig. 4. At both time points in both years, strong velocity jets at the anterior of the ascending aorta were seen for Patient 1, with maximum magnitudes reaching 0.7 and 1.0 [m s^{-1}] in 2022 and 2024, respectively; the magnitude in 2024 was overall greater than in 2022. Right-handed vortices were observed from the ascending aorta into the arch, and became stronger along the inner curvature before entering the descending aorta, in line with previous studies [23]. Higher velocities were seen in the descending aorta as a result of decreasing diameter. Systolic jets were predominantly centreline-aligned after the STJ, followed by generally smooth streamline trajectories across the arch and into the descending aorta. At T_{mid} , localised swirling was seen in both years in the ascending aorta due to flow deceleration, which extended into the arch and reorganised before entering the descending aorta.

For Patient 2, the pre-surgery case demonstrated a broader distribution of elevated velocity magnitudes in the ascending aorta with values up to 0.6 [m s^{-1}] at T_{sys} . The streamlines exhibited greater curvature compared with Patient 1, especially in the ascending aorta, with more pronounced secondary flow structures. Similar to Patient 1, right-handed vortical flow was observed before entering the descending aorta. Posterior views revealed swirling motion along the inner curvature of the arch, which extended into the proximal descending aorta. At T_{mid} , the reduction in velocity was accompanied by strong, chaotic secondary flows in the arch, the supra-aortic branches, and the descending aorta, with clusters of tightly curved streamlines. Post-surgery, the flow at peak systole was characterised by substantially uniform and concentrated velocity jets overall, with reduced vortical flow between the arch and the ascending aorta. However, a low-magnitude secondary flow was seen near the upper bulge of the distal aneurysm from the posterior view. The velocity magnitude nearly doubled, with peak values reaching up to 1.5 [m s^{-1}] at the kink of the graft in the ascending aorta. At T_{mid} , similar levels of local swirling to pre-surgical conditions were observed. Left-handed helical flow was present in the ascending aorta extending up to the arch, and an overall right-handed vortex was seen originating from the aneurysm through the descending aorta.

The helical flow observed from the velocity streamlines can be better visualised with the helicity density (H_k) iso-surfaces shown in Fig. 6. The helicity density was normalised with respect to the magnitude of vorticity and velocity; a magnitude of 1 thus indicates maximum helical flow. For both patients, the strongest helical flow was observed in the aortic arch, and more helicity patterns were overall present at T_{mid} . Small and discontinuous patches, mostly right-handed, were present in the ascending aorta and the arch at both follow-ups of Patient 1. Visibly greater spatial coverage of H_k was seen in 2024, particularly during

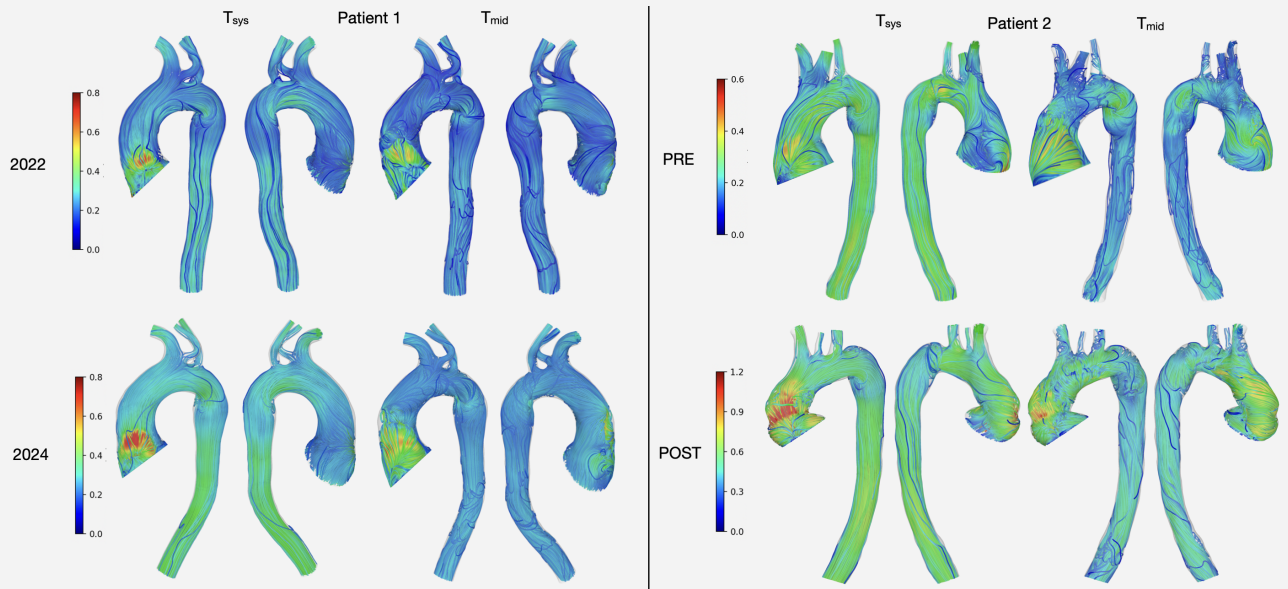


Figure 5: Velocity streamlines at peak systole (T_{sys}) and mid-systolic deceleration (T_{mid}). Note the difference in scale for each case.

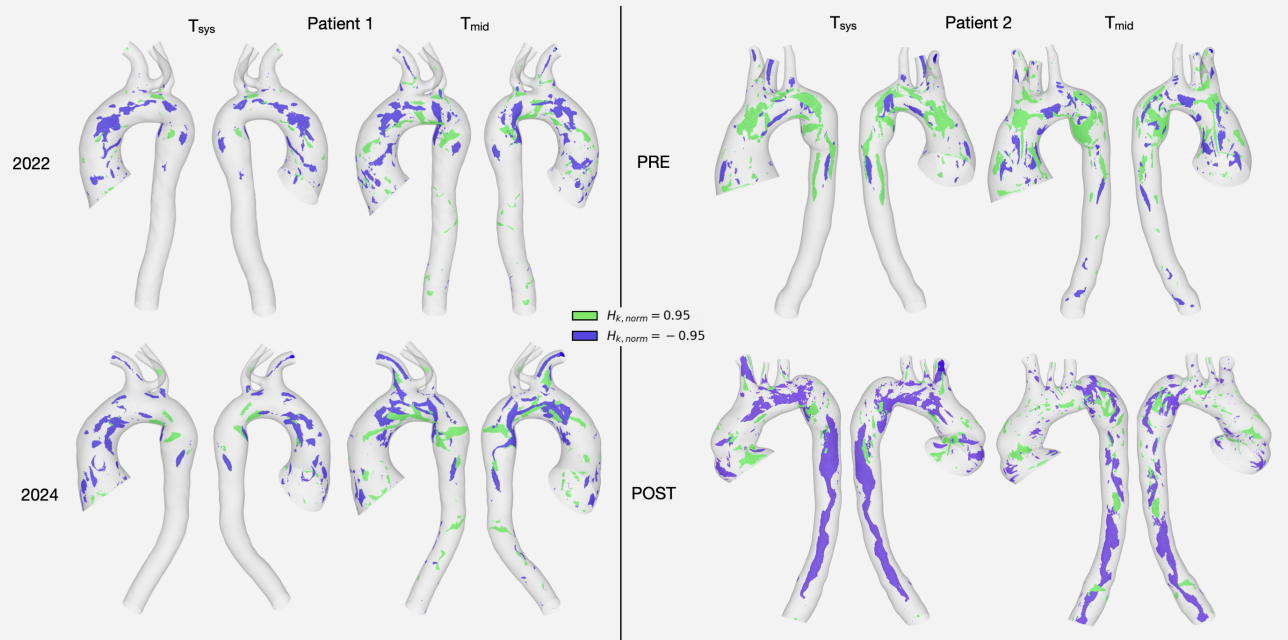


Figure 6: Normalised helicity densities iso-surfaces at 0.95 (green) and -0.95 (blue). Positive H_k values indicate right-handed helix, while negative H_k values indicate left-handed helix.

mid-systolic deceleration, with scattered patches extending into the supra-aortic branches and the proximal descending aorta.

In contrast, left-handed helical flow was more prominent in Patient 2 pre-surgery, present at the ascending aorta, and along both inner and outer curvatures of the aortic arch. These helicity patterns were largely preserved at T_{mid} , with additional strong helical flow in the region distal to the arch near the aneurysm. The helical patterns shifted dramatically post-surgery, dominated by left-handed helical patterns at T_{sys} . There was less coverage along the ascending aorta compared with the pre-operative state; significant helices were present in the arch, the brachycephalic artery, and descending aorta, corresponding to the velocity streamlines in Fig. 5. At T_{mid} , helical flow in the arch has reduced in

magnitude and reversed in direction.

3.3 Wall Shear Stress

Coloured maps of Time-Averaged Wall Shear Stress (TAWSS) are shown in Fig. 8. The following trend was consistent across all four cases: TAWSS was most distinct at the anterior of the ascending aorta, extending around the back and towards the arch. Moderately high values were found around the inner bend between the arch and the descending aorta due to a sudden change of geometry. Areas in red at the inlet had extreme values compared to the rest of the aorta; therefore, they were considered as a product of the inlet velocity boundary conditions from PC-MRI data, and thus neglected in calculations of the maximum value. TAWSS in Patient 1 reached 2.7 [Pa] in 2022 and 3.0

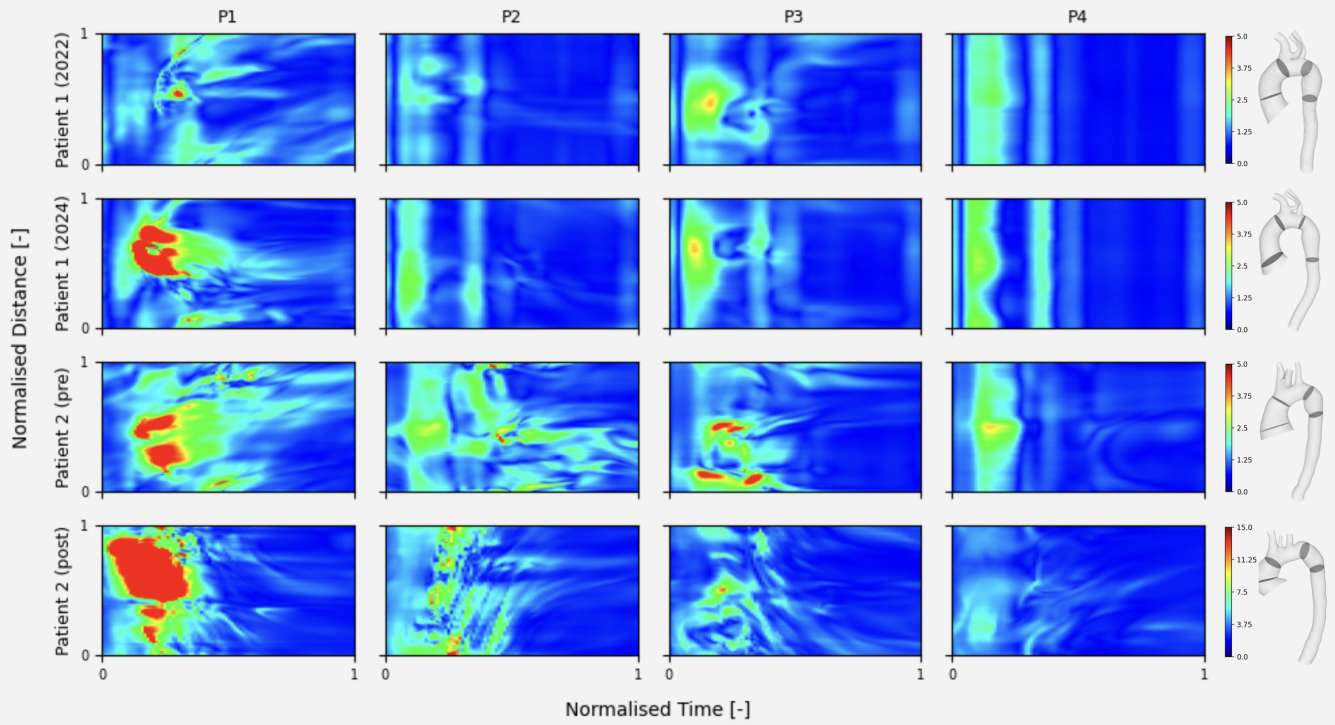


Figure 7: Temporal WSS [Pa] maps at four cross-sectional planes along the aorta over the cardiac cycle. The planes are unwrapped and mapped onto the axis normalised in distance, such that 0 and 1 correspond to the point that lies on the inner curvature of the aorta. Time is normalised with respect to the cardiac period. Note the difference in scale for Patient 2 Post.

[Pa] in 2024. In both years, these maximum values were found around the same area at the anterior of the ascending aorta; thus, this increase occurs simultaneously with further dilatation. Slight elevations were also found in the descending aorta in 2024, due to increased velocity as a result of the narrowed diameter. Aside from magnitude, this patient demonstrated virtually no difference in pattern across the two years.

The pre-surgery case for Patient 2 exhibited similar TAWSS to Patient 1, although with a broader distribution around the ascending aorta that extends into the arch. The maximum was also found on the sinus at 2.8 [Pa]. Magnitude around the distal aneurysm was also greater, with bands of high values above and below. After surgery, there was a significant increase throughout the entire aorta due to higher velocity. The maximum was observed at the kink of the graft, at 11.7 [Pa]. Patterns around the arch, the aneurysm, and the descending aorta were largely preserved.

To analyse differences in local WSS across time, four planes along the aorta were selected and unwrapped at the posterior and contours of WSS were plotted in the spatial and temporal axes in Fig. 7. The positions of these planes are: region immediately after the aorta valve (P1), before (P2) and after (P3) supra-aortic branches, and after the aneurysms (P4). Patient 1 demonstrated similar patterns in both years, with a more pronounced distribution in P1. For Patient 2, regions of high WSS at systole were seen dissipating sideways over time, but were generally in good agreement with Patient 1. The post-surgery images revealed extremely high and prolonged WSS across time in P1, which corresponds to the kink. P2 has shown a shift upwards (from front to back in an upwards motion) after

peak systole, possibly due to the left-handed helix observed previously.

Contours of oscillatory shear index (OSI) are shown in Fig. 9. For Patient 1, the spatial distribution of OSI was similar between 2022 and 2024. Patchy regions of elevated OSI were observed mainly around the arch

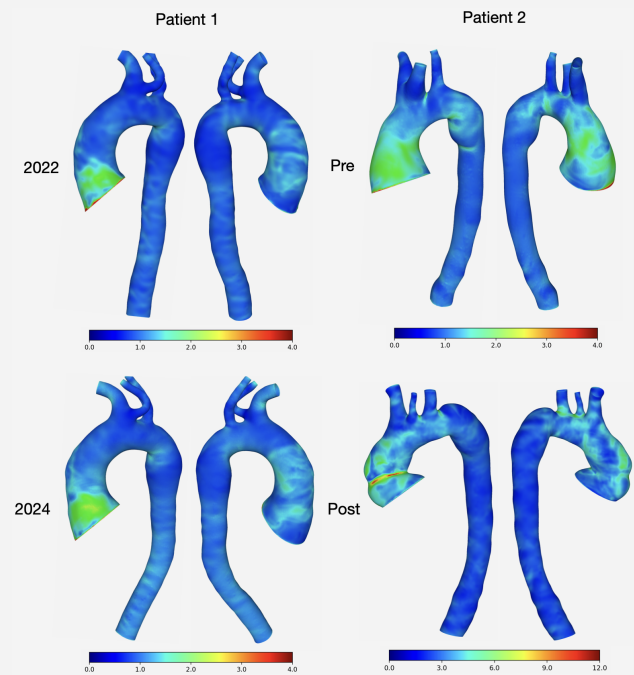


Figure 8: Coloured map of Time-averaged Wall Shear Stress (TAWSS) on aorta walls in [Pa]. Note the difference in scale for Patient 2 Post.

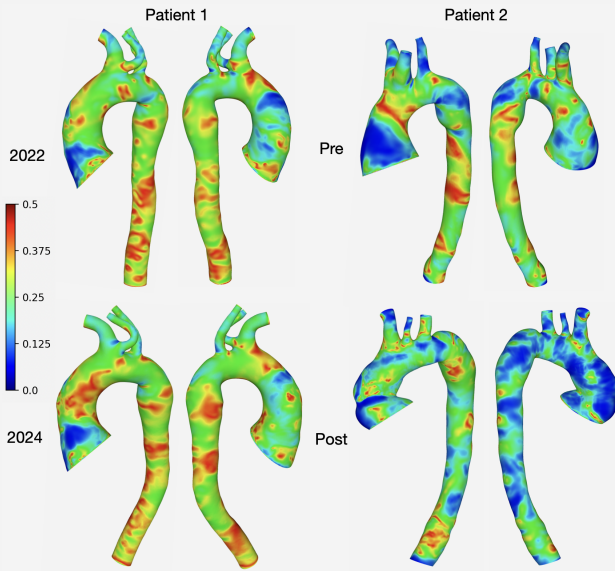


Figure 9: Oscillatory Shear Index (OSI) on aorta walls. A value of 0.5 reflects wall shear stress constantly change in direction across one cardiac cycle.

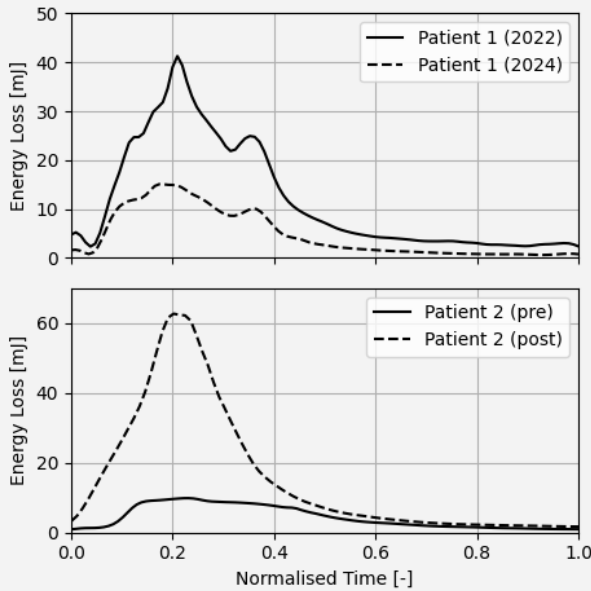


Figure 10: Energy loss in the entire aorta during a cardiac cycle. The time axis is normalised for better comparison.

and the descending aorta, with additional fluctuations in the ascending aorta in 2024. Before surgery, Patient 2 demonstrated similar patterns to Patient 1 overall; however, OSI in the sinus and supra-aortic branches was lower. After surgery, lower values of OSI were exhibited throughout, with generally uniform magnitudes.

3.4 Energy Loss

Energy loss through viscous dissipation across the entire fluid domain is compared in Fig. 10. The curves demonstrated a resemblance to the cardiac waveforms, mainly due to higher velocity gradients near the boundary region when inlet flow rates are high. Defined peaks corresponded to peak systole. Waveforms between 2022 and 2024 were

very similar for Patient 1, although a 60% decrease in total energy loss per cycle was present, from 11.8 to 4.8 [mJ]. On the other hand, there was a significant contrast between pre- and post-surgery waveforms for Patient 2. Peak energy loss was only at 9.9 [mJ] pre-surgery, comparable to Patient 1 in 2024. Post-surgery, total energy loss has increased by 350% from 4.4 to 15.96 [mJ], with peaks reaching 63 [mJ].

4 Discussion

Patients with Marfan syndrome have weakened connective tissue integrity; therefore, high pulsatile stresses and jet-like flow exacerbate dilatation. The aortic root is especially susceptible due to its complex wall architecture and flow patterns [24]. Both patients in this study had larger ascending aorta diameters than normal. Longitudinal analysis of the medically-treated patient demonstrated elevated WSS in the ascending aorta and higher OSI overall as the dilation progressed. This is in line with prior reports [25][26][27], confirming that the haemodynamic patterns of Patient 1 are characteristic of Marfan patients. The patient who underwent root replacement surgery showed similar trends to Patient 1 across all haemodynamic parameters analysed in this study before surgery. More vortex structures, broader distributions of WSS, and slightly lower magnitudes of TAWSS all indicate higher risk of aneurysm enlargement and dissection [14][11]. These patterns may be attributed to the relatively greater ascending aortic dilatation compared with the diameter of the arch. Thus, setting Patient 1 as a reference, the results supports that haemodynamic patterns of Patient 2 before surgery are also characteristic of Marfan patients, and postoperative results likely reflect surgical alterations superimposed on pre-existing features only.

Various abnormalities in haemodynamic parameters after aortic root replacement surgeries were found in this study. Compared to pre-surgical conditions and the patient who underwent conservative treatment, the velocity of blood flow has increased. Strong vortical and helical flow also appeared, particularly in the descending aorta. TAWSS also increased across the entire aorta, but there was less directional change of WSS during the cardiac cycle, as indicated by lower OSI values throughout.

These results are mostly consistent with existing literature on the risk of distal aortic diseases in Marfan patients, including aneurysm formation and type-B dissection, although some differences remain. Higher vortical flow was observed in 24 Marfan patients when compared to healthy controls and thus was associated with increased risk of disease progression in a 2012 study by Geiger et al. [9]. The study also highlighted the prominence of haemodynamics in the proximal descending aorta, as it is a frequent site of initial tear; the pathological shear forces acting on the vessel walls can be altered, affecting endothelial function and thus creating areas of increased risk [25][28]. In our study, the patient with the replaced root and ascending aorta has indeed developed type B dissection in 2020. The MR image is shown in Fig. 11. The initial tear lied on the inner curvature of the proximal descending aorta, a region where the vortex in the descending aorta first began to develop. Most studies associate elevated TAWSS with higher

dissection risks [14][26][8], which is consistent with the significant increase observed in this study. The discrepancy lies in the oscillatory shear index. Tajeddini et al. and Wen et al. [26][29] both associated dissection risk with increased OSI, however, results in this study were contradictory with lowered OSI values for the post-surgery case. One possible explanation is, since the velocity magnitude has nearly doubled, WSS becomes more unidirectional due to a higher axial momentum. The near wall shear vector points more consistently downstream for a larger fraction of the cardiac cycle as the forward jet is strengthened.



Figure 11: Aorta MR image of Patient 2 in 2020. The initial tear is located on the inner curvature of the proximal descending aorta (indicated at the colour transition).

It should be noted that the primary aim of aortic root and ascending aorta replacement was to limit further dilation of the ascending aorta, thereby preventing potential dissection. However, as shown by the significantly different haemodynamic parameters for Patient 2 between pre- and post-surgery, risks of distal aneurysm rupture and descending aorta dissection increased after surgery. Considering the fact that the patient with aortic root replacement suffered dissection three years after surgery, it can be concluded that signs of adverse vascular remodelling effects could be identified at an earlier stage. Kinks in the surgical graft used in replacement are rather rare occurrences [30], however, since the haemodynamic metrics exhibited by this patient shared similar patterns with dissection cases observed in other studies, this convergence suggests a justification for the results obtained.

The transitional behaviour of blood flow in Patient 2 was neglected in this study for consistency, since the other 3 cases were well within the laminar region. Xu et al. [31] compared Laminar and Large-Eddy Simulations (LES) in three dilated patient aortas, and found that no differences exist between large-scale parameters, with laminar simulations under-predicting TAWSS by only 5%. Similarly, Manchester et al. [32] reported small differences in viscous dissipation between well-defined laminar and LES simulations of up to 11%. Hence, there is less demand for implementing a turbulence model, such as LES or RANS-based models [33], when the primary aim is to compare haemodynamic metrics between treatment strategies. Nonetheless, this study could have benefited from using a turbulence model to improve the accuracy of absolute values of haemodynamic parameters. Peak Reynolds number is often calculated from the mean velocity

at a particular cross-section in the aorta at peak systole [34][35], which was followed in this study for comparison with a well-established stability criterion. However, were the local velocity magnitudes considered in the calculation, the Reynolds number at regions near the centre would lie well above the stability criteria. In a 2020 study by Saqr et al. [36], chaos theory was combined with hydrodynamic theory to conclude that pulsatile blood flow is in global hydrodynamic instability. Local turbulence effects are largely present, and the transition from laminar to turbulent flow is less defined. Hence, more attention should be given to improving the characterisation of flow regimes and to assessing whether a laminar model for blood flow is always oversimplified.

Few sources have studied energy loss in Marfan syndrome aortas using CFD, and most focused on turbulence dissipation [34]. However, as shown by Qiao et al. [18], energy lost in the aorta through turbulence dissipation accounts for only 5-10% of total energy loss, while viscous dissipation accounts for 50-70%, with the remaining energy lost to mutual work between blood flow and the wall. After surgery, Patient 2 saw a surge in energy loss through viscous dissipation. Since only Patient 2 post-surgery was in the turbulence regime, laminar results obtained in this study likely underestimated absolute post-surgery energy losses; thus the observed increase in viscous dissipation was conservative. Viscous dissipation is associated with frictional forces between fluid particles; hence, the increase could be explained by strong shear layers generated by the high velocity jet from the kinked graft. The implications of elevated energy loss are a higher afterload that the heart must overcome to pump blood into the aorta, although any correlation between this and risks of distal diseases would only be speculative.

A major limitation of this study was the use of the rigid-wall assumption. Aortic compliance, the ability of the aorta to expand and recoil with each cardiac cycle [37], was neglected due to the additional technical complexity of fluid-structure interaction modelling and the lack of patient-specific wall mechanical properties. Prior studies have shown that rigid-wall CFD often overestimates velocity magnitude and wall shear stress relative to compliance (e.g., linear elastic) models [38][39]. Accordingly, simulation accuracy could be improved by employing a compliance model. Nevertheless, these drawbacks did not compromise the comparative aims of our analysis since all cases were simulated under the same assumption, which preserved the validity of comparisons across patients and time-points, albeit with the caveat that absolute values might be biased upward.

5 Conclusion

In this study, a detailed assessment of the flow patterns in aortas of two Marfan patients that underwent different treatment methods was performed. CFD simulations were based on patient-specific geometries and boundary conditions obtained from PC-MRI data. Velocity streamlines, helicity density, WSS, OSI, and viscous energy loss were compared to make qualitative judgements on the risk of distal aortic diseases such as thoracic aneurysms and type-B dissection.

The results showed that while conservative treatment is effective and preserves haemodynamic patterns over time, it did not stop dilation and is only a temporary measure. Root and ascending aorta replacement surgeries are effective in mitigating short-term dissection risks involved with a largely dilated ascending aorta, but this kind of invasive procedure generates downstream implications, especially in the proximal and distal descending aorta. Thus, clinical decisions should be better informed by preliminary modelling before aorta geometry is altered. Future work should focus on incorporating turbulence and transition modelling to improve accuracy and on achieving higher resolution of flow regimes. Future studies focusing on a large cohort of patients are necessary to establish statistical significance and further the clinical relevance of the results.

Acknowledgements

The authors would like to thank Xiaoxin Kan and Binghuan Li for MR image acquisition and geometry reconstruction, and for sharing their technical expertise on ANSYS softwares. Special thanks to Chl  e Armour who kindly provided guidance on meshing.

References

- [1] A Dawson, Y Li, Y Li, P Ren, HG Vasquez, C Zhang, et al. "Single-Cell Analysis of Aneurysmal Aortic Tissue in Patients with Marfan Syndrome Reveals Dysfunctional TGF- β Signaling". In: *Genes* 13.1 (2021), p. 95.
- [2] DP Judge and HC Dietz. "Marfan's syndrome". In: *The Lancet* 366.9501 (2005), pp. 1965–1976.
- [3] SY Jang, Y Shin, J Huh, DK Kim, and TK Park. "Epidemiology of Marfan syndrome: a retrospective cohort study based on the Korean population". In: *Precis Future Med* 8.4 (2024), pp. 156–165.
- [4] EM Isselbacher, O Preventza, I Hamilton J. H., JG Augoustides, AW Beck, MA Bolen, et al. "2022 ACC/AHA guideline for the diagnosis and management of aortic disease: A report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines". In: *The Journal of Thoracic and Cardiovascular Surgery* 166.5 (2023), e182–e331.
- [5] SD Singh, XY Xu, NB Wood, JR Pepper, C Izgi, T Treasure, et al. "Aortic flow patterns before and after personalised external aortic root support implantation in Marfan patients". In: *Journal of Biomechanics* 49.1 (2016), pp. 100–111.
- [6] J Lotz, C Meier, A Leppert, and M Galanski. "Cardiovascular flow measurement with phase-contrast MR imaging: Basic facts and implementation". In: *Radiographics* 22.3 (2002), pp. 651–671.
- [7] CH Armour, B Guo, S Pirola, S Saitta, YF Liu, ZH Dong, et al. "The influence of inlet velocity profile on predicted flow in type B aortic dissection". In: *Biomechanics and Modeling in Mechanobiology* 20.2 (2021), pp. 481–490.
- [8] MM van Andel, P van Ooij, L Gottwald, V de Waard, AH Zwinderman, BJM Mulder, et al. "Abnormal aortic hemodynamics at predilection sites for dissection in patients with Marfan Syndrome: a 4D flow study". In: *European Heart Journal - Cardiovascular Imaging* 22.Supplement_2 (2021), jeab090.067.
- [9] J Geiger, M Markl, L Herzer, D Hirtler, F Loeffelbein, B Stiller, et al. "Aortic flow patterns in patients with Marfan syndrome assessed by flow-sensitive four-dimensional MRI". In: *Journal of Magnetic Resonance Imaging* 35.3 (2012), pp. 594–600.
- [10] A Guala, L Dux-Santoy, G Teixido-Tura, A Ruiz-Mu  oz, L Galian-Gay, ML Servato, et al. "Wall Shear Stress Predicts Aortic Dilation in Patients With Bicuspid Aortic Valve". In: *JACC: Cardiovascular Imaging* 15.1 (2022), pp. 46–56.
- [11] S Pirola, B Guo, C Menichini, S Simone, W Fu, Z Dong, et al. "4-D Flow MRI-Based Computational Analysis of Blood Flow in Patient-Specific Aortic Dissection". In: *IEEE Transactions on Biomedical Engineering* 66.12 (2019), pp. 3411–3419.
- [12] F Hellmeier, S Nordmeyer, P Yevtushenko, J Bruening, F Berger, T Kuehne, et al. "Hemodynamic Evaluation of a Biological and Mechanical Aortic Valve Prosthesis Using Patient-Specific MRI-Based CFD". In: *Artificial Organs* 42.1 (2018), pp. 49–57.
- [13] E Weigang, FA Kari, F Beyersdorf, M Luehr, CD Etz, A Frydrychowicz, et al. "Flow-sensitive four-dimensional magnetic resonance imaging: flow patterns in ascending aortic aneurysms". In: *European Journal of Cardio-Thoracic Surgery* 34.1 (2008), pp. 11–16.
- [14] J Geiger, R Arnold, L Herzer, D Hirtler, Z Stankovic, M Russe, et al. "Aortic wall shear stress in Marfan syndrome". In: *Magnetic Resonance in Medicine* 70.4 (2013), pp. 1137–1144.
- [15] MM van Andel, P van Ooij, V de Waard, LM Gottwald, RRJ van Kimmenade, AJ Scholte, et al. "Abnormal aortic hemodynamics are associated with risk factors for aortic complications in patients with Marfan syndrome". In: *IJC Heart & Vasculture* 43 (2022), p. 101128.
- [16] A Frydrychowicz, M Markl, D Hirtler, A Harloff, C Schlensak, J Geiger, et al. "Aortic Hemodynamics in Patients With and Without Repair of Aortic Coarctation: In Vivo Analysis by 4D Flow-Sensitive Magnetic Resonance Imaging". In: *Investigative Radiology* 46.5 (2011), pp. 317–325.
- [17] PM Vanhoutte, H Shimokawa, EHC Tang, and M Feletou. "Endothelial dysfunction and vascular disease". In: *Acta Physiologica* 196.2 (2009), pp. 193–222.
- [18] Y Qiao, J Fan, and K Luo. "Mechanism of blood flow energy loss in real healthy aorta using computational fluid-structure interaction framework". In: *International Journal of Engineering Science* 192 (2023), p. 103939.

- [19] A Leuprecht and K Perktold. "Computer Simulation of Non-Newtonian Effects on Blood Flow in Large Arteries". In: *Computer Methods in Biomechanics and Biomedical Engineering* 4.2 (2001), pp. 149–163.
- [20] S Pirola, Z Cheng, OA Jarra, DP O'Regan, JR Pepper, T Athanasiou, et al. "On the choice of outlet boundary conditions for patient-specific analysis of aortic flow using computational fluid dynamics". In: *Journal of Biomechanics* 60 (2017), pp. 15–21.
- [21] CA Kouser, NB Wood, WA Seed, R Torii, D O'Regan, and XY Xu. "A Numerical Study of Aortic Flow Stability and Comparison With In Vivo Flow Measurements". In: *Journal of Biomechanical Engineering* 135.1 (2012), p. 011003.
- [22] DG Dritschel. "A further twist to helicity". In: *Journal of Fluid Mechanics* 1010 (2025), F1.
- [23] PJ Kilner, GZ Yang, RH Mohiaddin, DN Firmin, and DB Longmore. "Helical and retrograde secondary flow patterns in the aortic arch studied by three-directional magnetic resonance velocity mapping". In: *Circulation* 88.5 (1993), pp. 2235–2247.
- [24] RE Pyeritz. "Etiology and pathogenesis of the Marfan syndrome: current understanding". In: *Annals of Cardiothoracic Surgery* 6.6 (2017).
- [25] A Frydrychowicz, R Arnold, D Hirtler, C Schlensak, AF Stalder, J Hennig, et al. "Multidirectional flow analysis by cardiovascular magnetic resonance in aneurysm development following repair of aortic coarctation". In: *Journal of Cardiovascular Magnetic Resonance* 10.1 (2008), p. 30.
- [26] F Tajeddini, DA Romero, YX Huang, TE David, M Ouzounian, CH Amon, et al. "Type B aortic dissection in Marfan patients after the David procedure: Insights from patient-specific simulation". In: *JTCVS Open* 21 (2024), pp. 1–16.
- [27] H.-H Wang, H.-H Chiu, W.-YI Tseng, and H.-H Peng. "Does altered aortic flow in Marfan syndrome relate to aortic root dilatation?" In: *Journal of Magnetic Resonance Imaging* 44.2 (2016), pp. 500–508.
- [28] AM Malek, SL Alper, and S Izumo. "Hemodynamic Shear Stress and Its Role in Atherosclerosis". In: *JAMA* 282.21 (1999), pp. 2035–2042.
- [29] J Wen, H Huang, Z Su, L Jiang, Q Gao, X Chen, et al. "Predicting the Risk of Type B Aortic Dissection Using Hemodynamic Parameters in Aortic Arches: A Comparative Study between Healthy and Repaired Aortas". In: *Computer Methods and Programs in Biomedicine* 230 (2023), p. 107326.
- [30] E Likaj, A Ibrahim, S Kuci, and J Zeitani. "Dacron® Graft Kinking Following Ascending Aorta Replacement Is Not Only a Cosmetic Issue". In: *Brazilian Journal of Cardiovascular Surgery* 40.2 (2025).
- [31] L Xu, T Yang, L Yin, Y Kong, Y Vassilevski, and F Liang. "Numerical Simulation of Blood Flow in Aorta with Dilation: A Comparison between Laminar and LES Modeling Methods". In: *Computer Modeling in Engineering & Sciences* 124.2 (2020), pp. 509–526.
- [32] EL Manchester, S Pirola, MY Salmasi, DP O'Regan, T Athanasiou, and XY Xu. "Evaluation of Computational Methodologies for Accurate Prediction of Wall Shear Stress and Turbulence Parameters in a Patient-Specific Aorta". In: *Frontiers in Bioengineering and Biotechnology* 10 (2022), p. 836611.
- [33] D Lopes, H Puga, J Teixeira, R Lima, J Grilo, J Dueñas-Pamplona, et al. "Comparison of RANS and LES turbulent flow models in a real stenosis". In: *International Journal of Heat and Fluid Flow* 107 (2024), p. 109340.
- [34] EL Manchester, S Pirola, MY Salmasi, DP O'Regan, T Athanasiou, and XY Xu. "Analysis of Turbulence Effects in a Patient-Specific Aorta with Aortic Valve Stenosis". In: *Cardiovascular Engineering and Technology* 12.4 (2021), pp. 438–453.
- [35] AF Stalder, A Frydrychowicz, MF Russe, JG Kovvink, J Hennig, K Li, et al. "Assessment of flow instabilities in the healthy aorta using flow-sensitive MRI". In: *Journal of Magnetic Resonance Imaging* 33.4 (2011), pp. 839–846.
- [36] KM Saqr, S Tupin, S Rashad, T Endo, K Niizuma, T Tominaga, et al. "Physiologic blood flow is turbulent". In: *Scientific Reports* 10.1 (2020), p. 15492.
- [37] RA Ramdax, A Matar, CA Valdes, S Patel, B Tadesse, MI Chen, et al. "Aortic Compliance: A Narrative Review". In: *J Endovasc Ther.* (2025).
- [38] C Reeps, M Gee, A Maier, M Gurdan, HH Eckstein, and WA Wall. "The impact of model assumptions on results of computational mechanics in abdominal aortic aneurysm". In: *Journal of Vascular Surgery* 51.3 (2010), pp. 679–688.
- [39] M Bonfanti, S Balabani, JP Greenwood, S Puppala, S Homer-Vanniasinkam, and V Díaz-Zuccarini. "Computational tools for clinical support: a multi-scale compliant model for haemodynamic simulations in an aortic dissection based on multi-modal imaging data". In: *Journal of The Royal Society Interface* 14.136 (2017).
- [40] BA Craven, EG Paterson, GS Settles, and MJ Michael. "Development and Verification of a High-Fidelity Computational Fluid Dynamics Model of Canine Nasal Airflow". In: *Journal of Biomechanical Engineering* 131.9 (2009), p. 091002.

Numerical Modelling and Process Evaluation of Structural Transitional Adsorbents for Adsorption-based Carbon Capture

Ross Lloyd and Eric Jayawant

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

As global CO₂ concentrations continue to rise, carbon-capture technologies are increasingly important for their role in reducing greenhouse gas emissions and mitigating global warming. Structural transitional adsorbents (STAs) have become a particular area of interest due to their unique ability to undergo phase change. This class of metal organic frameworks (MOFs) are an attractive candidate for adsorption-based carbon capture as this phase change results in rapid adsorption of CO₂ over a small pressure range, significantly increasing the selectivity of the material. Despite this, few studies have attempted to assess the performance of these materials in the context of process design. This paper presents the numerical modelling of STAs within conventional cyclical adsorption cycles, realising their potential for industrial application. Through simulations, STA materials explored were found to outperformed traditional materials under pressure swing adsorption (PSA) and vacuum swing adsorption (VSA), with STAs offering a significant energy reduction under PSA application without compromising key performance indicators (KPIs) such as purity and recovery. Pressure vacuum swing adsorption (PVSA) found that individual STAs delivered both greater working capacities and energy reduction at the same operating conditions, demonstrating that the integration of these materials would significantly improve the efficiency and economics of carbon capture processes.

Keywords: carbon capture, adsorption, structural transitional adsorbents, isotherm modelling, process performance, KPIs

1. Introduction

Since the 1960s, atmospheric CO₂ concentrations have risen from ~320 ppm to ~420 ppm by 2024, primarily driven by the energy and manufacturing industries as a result of population growth.¹ This increase has accelerated global warming and ocean acidification, triggering widespread environmental and ecological impacts.² This anthropogenic release of CO₂ is mostly the result of an overreliance on fossil fuels, which accounted for approximately 86.6% of global energy supply in 2024³, with coal being the most carbon-intensive of these sources releasing 740 – 1689 g CO₂ per kWh.⁴

Despite government policies (such as the Paris agreement⁵) aimed at reducing CO₂ emission, fossil-fuel consumption continues to climb as renewable energy sources struggle to meet rising global demand, highlighting the need for alternative strategies while sustainable energy infrastructure is scaled up.

Post-combustion carbon capture (PCCC) is one of the most mature and widely implemented industrial approaches to reduce emission, using amine-based solvents which face high solvent cost, environmental and safety concerns, and significant energy requirements for solvent regeneration⁶, contributing to substantial operational expenses, often rendering this method economically unfavourable as energy consumption can lie anywhere from 20%-30% of a power plants output⁷.

As a result, alternative separation methods are being investigated, with solid adsorption emerging as a promising option, currently being tested at pilot scale with encouraging results⁸, alongside ongoing screening of potential materials for solid-based carbon capture.

A new class of adsorbent materials known as structural transitional adsorbents, a type of metal organic framework capable of undergoing a phase change once certain process conditions are met, show promise for this application. These STAs could potentially offer a greater energy reduction and greater performance when compared to traditional materials, as this structural change provides the STA with a greater species selectivity.

Despite having detailed models to describe STA behaviour, minimal implementation of these flexible MOFs has been performed in a process setting, largely due to the complexity when integrating unique equilibrium adsorption behaviour into process models.

This paper aims to evaluate the performance of STA materials in an industrial application, implementing them into a cyclic adsorption process to determine whether these materials can improve process performance when compared to traditional adsorbents.

2. Background

2.1. Cyclic Adsorption Process

Industrial adsorption is a cyclical process comprising of 4 stages, with Figure 1 highlighting this. The blowdown stage ($\alpha \rightarrow \beta$) reduces the pressure to an intermediate level to desorb an undesired product from the bed. This is then followed by the evacuation stage ($\beta \rightarrow \gamma$) where the pressure is reduced further to desorb the desired species (in this case CO₂). The feed pressurisation (FP) stage ($\gamma \rightarrow \delta$) follows the evacuation stage and uses a fraction of the feed gas to re-pressurise the vessel and ensure that the target component is adsorbed in the Adsorption stage ($\delta \rightarrow \alpha$).

Traditionally, this adsorption process is operated under three distinct ways:

- PSA: where the low pressure is atmospheric conditions and the high pressure is above atmospheric.
- VSA: where the low pressure is under vacuum and the high pressure is at atmospheric conditions.
- PVSA: where the high pressure is above atmospheric conditions and the low pressure is under vacuum.

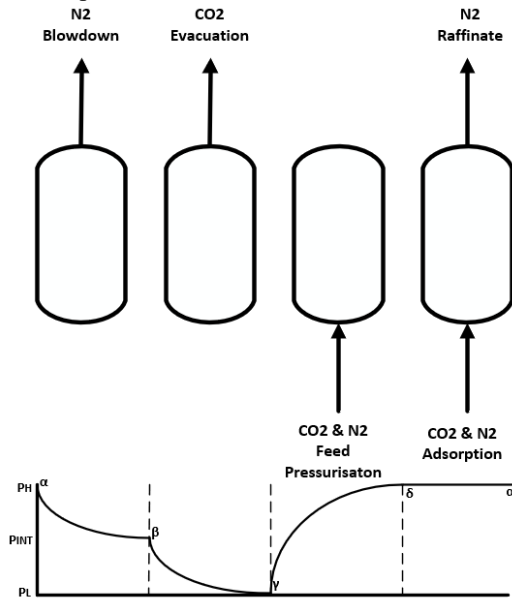


Figure 1. Schematic of 4 stage cyclical adsorption cycle with feed pressurisation.

2.2. Traditional Adsorbents

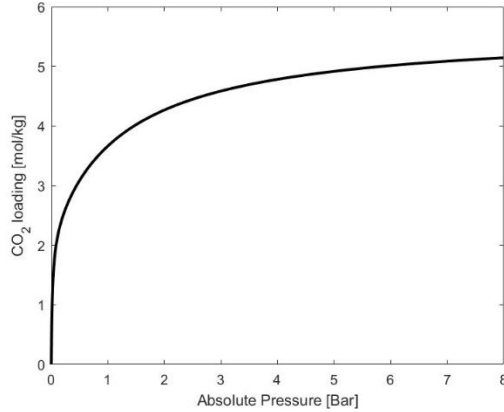


Figure 2. Schematic of traditional adsorption isotherm

Traditional adsorbents present the following concave and continuous equilibrium isotherm illustrated in figure 2, whose rigid pores contain active sites that enable physical and/or chemical binding of species to the material, determined by intermolecular interactions of the adsorbent and gases involved⁹.

In the context of carbon capture, these materials are mature and well understood, with studies such as Ward, A. et al. 2023 having screened

and optimised binary combinations of traditional materials in process settings.¹⁰

2.3. Structural Transitional Adsorbents

STAs differ from traditional materials by exhibiting a rapid adsorption step.¹¹ This behaviour is a result of the interactions between the adsorbate gases and the framework species, where a phase change occurs dependant on gas composition, temperature and pressure, expanding from a narrow pore (NP) state to a large pore (LP) state. These larger pores can accommodate a greater number of molecules within the structure of the STA and is often more selective for specific adsorbates. These features increase the working capacity within specific pressure ranges, making STAs particularly attractive for gas separation applications⁹.

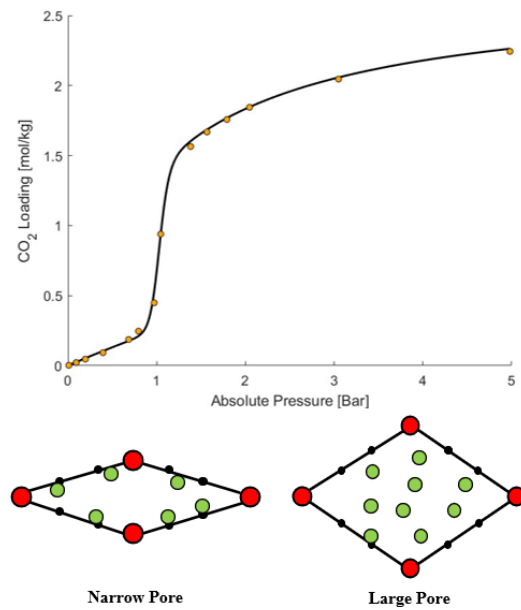


Figure 3. Schematic of STA adsorption isotherm and phase change phenomena

3. Methods

3.1. Modelling of STA isotherms

Hiraide-et-al.(2023)¹² have developed the new theoretical isotherm equations for STA materials which have two phases, modelling both narrow and large pore sites with non-competitive Langmuir adsorption for each state. The total adsorption is constructed by weighting the adsorption at each phase according to what fraction of pores have undergone this structural transition at a specific temperature, pressure and gas composition.

$$K_{\alpha,i} = K_{\alpha,i}^0 \exp\left(\frac{Q_a}{RT}\right) \quad (1)$$

Here, $K_{\alpha,i}$ is the Langmuir adsorption constant for phase α (NP or LP), and species i at temperature T . $K_{\alpha,i}^0$ is a constant, Q_a is the heat of adsorption and R is the gas constant.

Adsorption at phase α can then be calculated via:

$$q_i^\alpha = q_{\alpha,i}^{max} \frac{K_{\alpha,i} P y_i}{1 + K_{\alpha,i} P y_i} \quad (2)$$

Where $q_{\alpha,i}^{max}$ is a constant, P is the absolute pressure and y_i is the mole fraction of species i .

The total adsorption is then given by:

$$q_i^* = (1 - \sigma) q_i^{NP} + \sigma q_i^{LP} \quad (3)$$

Where q_i^* is the total adsorption of species i on the STA, and σ is a sigmoid function, representing the fraction of pores in the LP state. This sigmoid function is calculated with a material specific exponent, s , and the dimensionless variable $y_{val,i}$, which combines the adsorption contributions of each state with the free-energy difference between NP and LP phases:

$$\sigma = \frac{y_{val,i}^s}{1 + y_{val,i}^s} \quad (4)$$

$$y_{val,i} = \frac{(1 + K_{LP,i} P y_i)^{q_{LP,i}^{max}}}{(1 + K_{NP,i} P y_i)^{q_{NP,i}^{max}}} \exp\left(-\frac{\Delta U_{host} - T \Delta S_{host}}{RT}\right) \quad (5)$$

3.2. Modelling of Traditional Isotherms

To maintain consistency with the non-competitive STA model, traditional materials have been modelled using a non-competitive dual-site Langmuir (DSL) isotherm to ensure fair assessment of KPIs between traditional and STA adsorbents

This model accounts for adsorption of two types of sites within the porous structure, each with distinct capacities. The DSL model can accurately capture the adsorption behaviour of traditional materials¹³, expressed as:

$$q_i^* = q_{1,i}^{max} \frac{K_{1,i} C_i}{1 + K_{1,i} C_i} + q_{2,i}^{max} \frac{K_{2,i} C_i}{1 + K_{2,i} C_i} \quad (6)$$

Where $q_{1,i}^{max}$ and $q_{2,i}^{max}$ are the saturation capacities of sites 1 and 2 of the material. $K_{1,i}$ and $K_{2,i}$ are the Langmuir constants at temperature T , for species i at sites 1 & 2 respectively. C_i is the gas phase concentration of species i , calculated via the ideal gas law:

$$C_i = \frac{P y_i}{RT} \quad (7)$$

The temperature dependence of the Langmuir constants follow the same form as the STA model (Equations. (1) & (2)), with site specific pre-exponential constants and heat of adsorption.

3.3. Process Modelling

Modelling of industrial cyclical adsorption processes are both mathematically and computationally intensive, with real processes exhibiting a multitude of condition variability involving temperature, pressure and concentration gradients, alongside resistances to mass transfer and dispersion.

To assess the performance of adsorbents, the batch analogue adsorption model (BAAM), developed by Balashankar et al(2019), was selected to screen adsorbent materials, the model relies on the following assumptions:¹⁴

- Well-mixed and homogeneous column.
- Ideal fluid phase behaviour.
- Uniform adsorbent properties and bed voidage (ε) across column.
- No mass transfer resistances

Table 1. Material Parameters for STA and DSL adsorption isotherms

STA Parameters	F4MIL140		ZIF7	
Species	CO ₂	N ₂	CO ₂	N ₂
K_{NP}^0 [Bar ⁻¹]	2.05×10^{-6}	3.29×10^{-5}	5.60×10^{-7}	1.35×10^{-6}
Q_{NP} [J/mol]	3.09×10^4	1.84×10^4	3.29×10^4	2.06×10^4
q_{NP}^{max} [mol/kg]	2.68	0.98	1.68	1.68
K_{LP}^0 [Bar ⁻¹]	5.57×10^{-7}	3.29×10^{-5}	8.98×10^{-6}	3.36×10^{-5}
Q_{LP} [J/mol]	4.13×10^4	1.84×10^4	3.31×10^4	1.82×10^4
q_{LP}^{max} [mol/kg]	2.68	0.98	1.68	1.68
ΔU_{host} [J/mol]	4.93×10^3	4.93×10^3	4.95×10^3	4.95×10^3
ΔS_{host} [J•K ⁻¹ •mol ⁻¹]	4.57×10^{-8}	4.57×10^{-8}	2.58×10^{-8}	2.58×10^{-8}
s [kg/mol]	16.89	16.89	17.35	17.35
ρ [kg/m ³]	1720 ¹⁶		1520 ¹⁷	
DSL Parameters	Zeolite13X ¹⁰		Activated Carbon ¹⁰	
Species	CO ₂	N ₂	CO ₂	N ₂
K_1^0 [m ³ /mol]	1.71×10^{-7}	8.55×10^{-8}	8.07×10^{-7}	3.91×10^{-10}
Q_1 [J/mol]	3.35×10^4	2.45×10^4	2.00×10^4	2.82×10^4
q_1^{max} [mol/kg]	3.40	3.40	9.12	9.12
K_2^0 [m ³ /mol]	2.69×10^{-7}	8.55×10^{-8}	6.29×10^{-6}	1.23×10^{-5}
Q_2 [J/mol]	4.27×10^4	2.45×10^4	2.09×10^4	1.27×10^4
q_2^{max} [mol/kg]	2.22	2.22	2.69	2.69
ρ [kg/m ³]	1130		670	

- Isothermal operation.

These assumptions simplifying the adsorption column and remove much of the process variability associated with the process. A binary CO₂/N₂ feed was used, with a CO₂ composition of 15%, representative of coal-fired flue gas.¹⁵

Under the above assumptions, the overall and component mass balances for the column are:
Overall mass balance:

$$\frac{V\varepsilon}{RT} \frac{dP}{dt} + m \left(\frac{dq_{CO_2}^*}{dt} + \frac{dq_{N_2}^*}{dt} \right) = -Q \quad (8)$$

Component mass balance:

$$\frac{V\varepsilon}{RT} \frac{dPy_{CO_2}}{dt} + m \frac{dq_{CO_2}^*}{dt} = -Qy_{CO_2} \quad (9)$$

Where Q is the total molar flowrate leaving the column, m refers to the mass of adsorbent, and V is the total bed volume, calculated from adsorbent density, mass and bed voidage.

These equations can be combined to eliminate flowrate and time components, resulting in an ordinary differential equation (ODE) describing how gas phase composition (y_{CO_2}) changes with pressure at cyclical steady state, encompassing both the blowdown and evacuation steps ($\alpha \rightarrow \beta \rightarrow \gamma$).

$$\frac{dy_{CO_2}}{dP} = \frac{a_1 y_{CO_2} - a_2}{f_2 - f_1 y_{CO_2}} \quad (10)$$

Where,

$$a_1 = \frac{V\varepsilon}{RT} + m \left(\frac{\partial q_{CO_2}^*}{\partial P} + \frac{\partial q_{N_2}^*}{\partial P} \right) \quad (11)$$

$$a_2 = \frac{y_{CO_2} V\varepsilon}{RT} + m \frac{\partial q_{CO_2}^*}{\partial P} \quad (12)$$

$$f_1 = m \left(\frac{\partial q_{CO_2}^*}{\partial y_{CO_2}} + \frac{\partial q_{N_2}^*}{\partial y_{CO_2}} \right) \quad (13)$$

$$f_2 = \frac{PV\varepsilon}{RT} + m \frac{\partial q_{CO_2}^*}{\partial y_{CO_2}} \quad (14)$$

The mass of adsorbent was fixed at 1kg, and bed voidage was considered to be 0.37.

The feed pressurisation stage, ($\gamma \rightarrow \delta$), is later calculated by solving the total and component mass balances respectively:

$$\frac{(N_{CO_2, total}^\gamma + N_{N_2, total}^\gamma) + N_{FP}}{(N_{CO_2, total}^\delta + N_{N_2, total}^\delta)} = \quad (15)$$

$$N_{CO_2, total}^\gamma + N_{FP} \cdot y_{CO_2}^{feed} = N_{CO_2, total}^\delta \quad (16)$$

Where N_i refers to the total number of mols of species i at respective stages, N_{FP} and $N_{CO_2, total}^\delta$ are unknown variables.

The adsorption stage ($\delta \rightarrow \alpha$) assumed plug for the feed gas used to re-pressurise the column, breaking through only once the bed is fully saturated and the CO₂ composition of the bed is equal to that of the feed. The total and component mass balances are:

$$\frac{(N_{CO_2, total}^\delta + N_{N_2, total}^\delta) + N_{Feed}}{(N_{CO_2, total}^\alpha + N_{N_2, total}^\alpha)} = \quad (17)$$

$$N_{CO_2, total}^\delta + N_{Feed} \cdot y_{CO_2}^{feed} = N_{CO_2, total}^\alpha \quad (18)$$

Where N_{Feed} is the number of moles of feed gas required to bring the cycle back to a saturated state (α).

All equations were solved using MATLAB 2024b. Equation (10) was solved using ode15s with step size of 1 Pa from high pressure (P_H) to low pressure (P_L), while equations (15)-(18) are solved simultaneously utilising the Fsolve function to calculate unknown variables N_{FP} , N_{feed} and $N_{CO_2, total}^\delta$.

3.4. Numerical Methods

Equations (11)-(14) require the partial derivatives of CO₂ and N₂ adsorption with respect to both pressure and gas-phase composition. Due to the complexity of the STA isotherm, numerical methods were used to compute these partial derivatives required for BAAM, where computed using a central difference method:

$$\frac{\partial q^*}{\partial x} \approx \frac{q^*(x + \Delta x) - q^*(x - \Delta x)}{2\Delta x} \quad (19)$$

where $x = P$ or y_{CO_2} .

To accurately resolve the step associated with phase change, the following step sizes were used:

$$\Delta y_{CO_2} = 10^{-15}$$

$$\Delta P = 10^{-5} [Bar]$$

Larger step sizes have cause MATLAB's ODE solver to fail due to its inability to handle the rapid decrease in equilibrium adsorption, while smaller values lead to unrealistic desorption behaviour, with CO₂ composition increasing immediately to 1.

3.5. Key Performance Indicators

To assess the effectiveness of a cyclic adsorption cycle for solid-based carbon capture, key performance indicators are required. These KPIs allow for the comparison of process outputs and metrics when integrating different materials under distinct operating conditions.

The four KPIs considered in this analysis are:

1. CO₂ Purity (Pu_c): Where high purity is ideal.
2. CO₂ Recovery (Re_c): High recovery ensures lower CO₂ emissions.
3. Specific Energy (E_n): Energy consumption per tonne of CO₂ captured, which is related to operational costs.
4. Working Capacity (WC): The mass of CO₂ captured per m^3 adsorbent.

$$Pu_c [\%] = \frac{N_{CO_2}^{\beta \rightarrow \gamma}}{N_{N_2}^{\beta \rightarrow \gamma} + N_{CO_2}^{\beta \rightarrow \gamma}} * 100 \quad (20)$$

$$Re_c [\%] = \frac{N_{CO_2}^{\beta \rightarrow \gamma}}{N_{CO_2}^{\alpha}} * 100 \quad (21)$$

$$WC [Tonne CO_2 \cdot m^{-3}] = N_{CO_2}^{\beta \rightarrow \gamma} * M * \rho \quad (22)$$

$$En [kWh \cdot Tonne CO_2^{-1}] = \frac{W_{BLO} + W_{EVAC} + W_{ADS}}{N_{CO_2}^{\beta \rightarrow \gamma} * M} \quad (23)$$

Energy consumption is calculated as the sum of the work across all process steps:

$$W_{BLO} = \int_{N_{total}^{\alpha}}^{N_{total}^{\beta}} \frac{1}{\eta} \frac{k}{k-1} RT \left[\left(\frac{1}{P} \right)^{\frac{k-1}{k}} - 1 \right] dN \quad (24)$$

$$W_{EVAC} = \int_{N_{total}^{\beta}}^{N_{total}^{\gamma}} \frac{1}{\eta} \frac{k}{k-1} RT \left[\left(\frac{1}{P} \right)^{\frac{k-1}{k}} - 1 \right] dN \quad (25)$$

$$W_{ADS} = \int_{N_{total}^{\gamma}}^{N_{total}^{\alpha}} \frac{1}{\eta} \frac{k}{k-1} RT \left[\left(\frac{1}{P} \right)^{\frac{k-1}{k}} - 1 \right] dN \quad (26)$$

Here, an adiabatic constant (k) of 1.4 was assumed, while pump efficiency is a function of pressure. M refers to the molecular weight of CO₂ [tonnes $\cdot$ mol⁻¹]

Blowdown and evacuation work were neglected unless the pressure of the process was less than 1 Bar, as the vacuum pump was assumed to be inactive above atmospheric pressure. Similarly, compression work associated with adsorption & feed pressurisation was ignored if the high pressure was below or equal to 1 Bar.

5.6. Case Study Operating Conditions

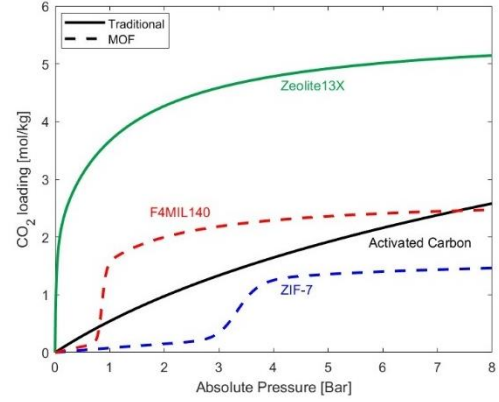


Figure 4: CO₂ isotherms of adsorbent materials at different absolute pressures with a CO₂ molar composition of 15%.

The performance of STA materials against traditional adsorbents will be evaluated under all three distinct ways this cyclical process is operated.

Table 2. Fixed operating conditions for cyclic adsorption case studies.

	P_H [Bar]	P_L [Bar]	Temperature [K]
VSA	1	0.1	298.15
PSA	8	1	298.15
PVSA	8	0.1	298.15

The selection of P_H was considered appropriate, as at this pressure, CO₂ adsorption increases negligibly with respect to pressure, with all materials considered at its maximum adsorption capacity. Pressures greater than 8 Bar would result in unnecessary energy consumption associated with compression.

P_L was chosen based on the lowest practical pressure when operating in industry, with 0.1 Bar balancing both energy consumption and pump performance.¹⁸

Traditional materials Zeolite-13X and Activated Carbon (AC) were selected for comparison against STA materials F4MIL-140 and ZIF-7. This set provides a broad range of CO₂ adsorption and two distinct phase-change pressures, allowing for the examination of how phase change influences STA material KPIs relative to traditional materials.

The decision to operate at standard reference temperature ensures the accuracy of our selected material isotherm parameters. Ensuring KPIs are reliable and comparable across all materials and operating modes.

5.7. Optimality

To assess the practicality of using different materials under the conditions tested, the following

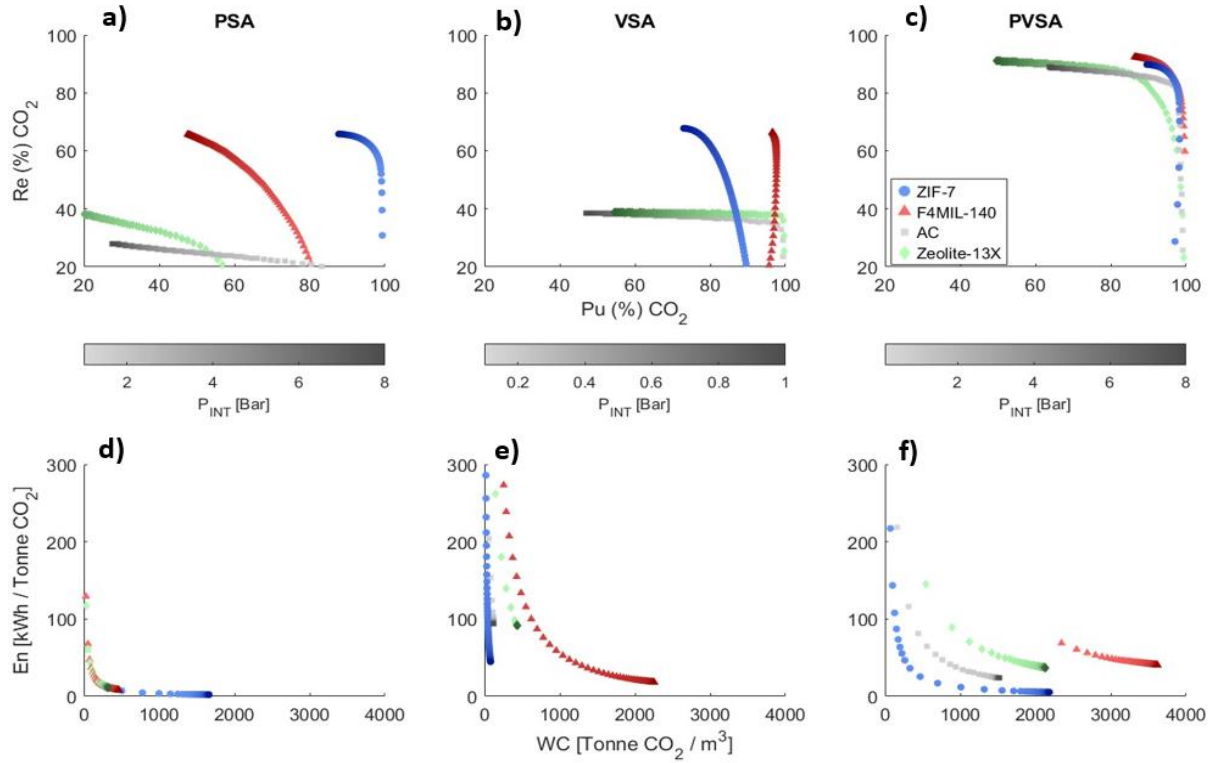


Figure 5: Performance of Traditional and STA materials. Panels (a)-(c) compare Purity CO₂ / Recovery CO₂ while panels (d)-(f) compare Specific Energy / Working Capacity at different intermediate pressures for respective cyclical operating conditions.

calculation was used to identify the optimal point where both purity and recovery are maximised:

$$\min \left[\sqrt{(100 - Pu_c)^2 + (100 - Re_c)^2} \right] \quad (27)$$

Despite maximising purity and recovery, this point will not necessarily represent the best operating condition if the resulting purity or recovery fails to meet process-specific targets. Typically, CO₂ separation processes require a minimum of 95% purity and recovery targets of 85-95%, with lower recoveries usually being more realistic and acceptable.^{19,20,21}

6. Results

For each of the four materials, the trade-off between purity and recovery is observed: Where recovery is greater at higher intermediate pressures, and purity is greater at lower intermediate pressures. Specific energy consumption follows the opposite trend, decreasing as intermediate pressures increase, while the working capacity decreases with decreasing intermediate pressure.

6.1 PSA

ZIF-7 achieves the best performance under PSA achieving purities greater than 90% and recoveries greater than 65%, having dominant KPIs amongst the materials studied. F4MIL-140 shows lower purity compared with activated carbon, however, it still outperforms traditional materials for recovery, working capacity and energy efficiency for these cycle conditions. STA materials are shown

to have substantially greater working capacities and reduced specific energy at this operating condition.

Table 3. Optimal KPIs for PSA cycle

PSA	P_{INT}	Pu	Re	WC	En
ZIF-7	6.1	90.9	65.3	164.9	2.2
F4MIL-140	4.6	59.7	57.7	41.2	8.8
AC	1.5	83.2	20.1	23.0	15.8
Zeolite-13x	1.8	50.8	27.2	21.5	16.9

6.2 VSA

Table 4: Optimal KPI's for VSA cycle

VSA	P_{INT}	Pu	Re	WC	En
ZIF-7	0.9	76.2	66.7	8.0	45.6
F4MIL-140	1.0	96.5	66.4	225.7	18.0
AC	0.2	94.4	35.7	11.8	101.3
Zeolite-13x	0.2	96.2	38.1	42.9	93.3

Simulations under VSA have found F4MIL-140 delivers the best overall performance, with a purity of 96.5% and recovery of 66.4%. Its greatest advantage is that, the working capacity is an order of magnitude higher than other materials, while still having the lowest specific energy. Despite recovery and specific energy being better for ZIF-7 than both traditional adsorbents, the purity of the evacuation stream shows it should not be used for this application. Additionally, energy demands under these conditions are substantially higher than

those for PSA, due to the work associated with operating under vacuum.

6.3 PVSA

Table 5: Optimal KPI's for PVSA cycle

PVSA	P_{INT}	P_u	Re	WC	En
ZIF7	5.1	93.5	88.6	215.5	5.50
F4MIL140	3.7	93.0	90.0	353.0	41.2
AC	1.7	93.6	84.4	144.3	24.9
Zeolite-13x	1.6	87.2	85.4	200.0	39.7

PVSA operation produces the highest purity-recovery combinations across all materials, with all adsorbents having a competitive purity and recovery, indicating they could all be viable in a process application. However, at the optimal trade-off points, both STA materials are advantageous when compared to traditional materials, having a higher working capacity or lower specific energy at the same operating point.

7. Discussion

The indirect relationship between purity and recovery that is exhibited across process conditions can be explained by the desorption paths taken by CO_2 and N_2 . At higher intermediate pressures, less CO_2 is lost during blowdown, meaning that the overall recovery will be higher. However, with a decrease in the intermediate pressure, purity increases as the gradient of the CO_2 isotherm is smaller than the N_2 isotherm as shown in Figure 7 (a)-(b). Evacuation where this is the case is 'wasted' operation from a purity standpoint since CO_2 is mostly being retained in the adsorbent material along the relatively flat gradient section. To maximise purity, ideal intermediate pressure would be placed just before the sharp decrease in loading. At this point, the decrease in CO_2 loading is always more rapid than the decrease in N_2 loading.

The total energy required for each material during each case study remains constant, as it depends on the initial and final pressure of the cycle. Consequently, greater recovery reduces the specific energy of the process as more CO_2 is present within

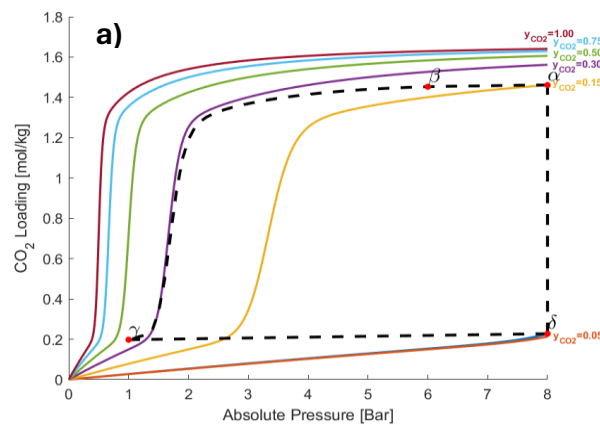


Figure 7: panel (a) shows the CO_2 desorption profile of ZIF-7 for PSA and panel (b) shows the N_2 desorption profile of ZIF-7 for PSA.

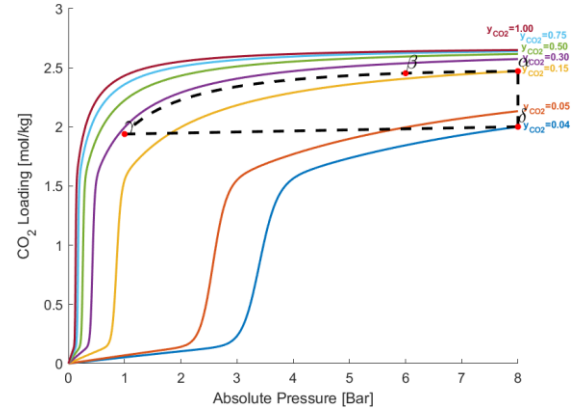


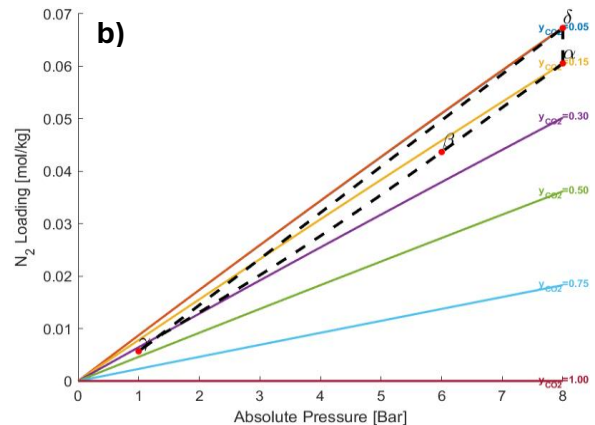
Figure 6: CO_2 desorption profile of F4MIL-140 for PSA

the evacuation stage for the same total energy, which is associated with high pressure. Similarly, because the mass of adsorbent within the bed does not change, higher recoveries result in greater working capacities.

7.1 PSA

Under these conditions, lower purities for F4MIL-140 reflect the fact that phase transition does not occur until vacuum conditions (Figure 6), preventing the large extraction of CO_2 that would occur upon desorption from being exploited. This explains why the working capacity is lower for F4MIL-140 compared to ZIF-7, where the phase change phenomena is utilised. Consequently, the PSA cycle is only practical for ZIF-7, yet the recoveries for this material could be improved with alternative cycle conditions.

Equilibrium isotherms (Figure 4) show that to achieve greater purities for traditional materials pressure must be dropped into vacuum conditions. One of the major drawbacks of a relatively constant loading gradient is displayed in the results for activated carbon which shows very low recoveries, the poorest out of the 4 materials. It is important to notice that recoveries and purities are not optimal for these cycle conditions, even for ZIF-7, since there is



further evacuation of CO₂ that is not taken advantage of.

7.2 VSA

The flip between purity-recovery trends for STA materials ZIF-7 and F4MIL-140 arises due to the location of pore transition associated with the STAs, with F4MIL-140 being more suited to vacuum conditions since its phase change occurs below 1 bar while ZIF-7 does not undergo phase change. Because the large desorption of CO₂ is taken advantage of, both working capacity and purity increase substantially.

Additionally, ZIF-7 purity decreases for the VSA cycle when compared to the PSA cycle. This is expected since the CO₂ isotherm gradient at these conditions (at pressures below the phase change) is much smaller than in the PSA cycle.

Overall, the only material that that is suitable for this cycle is F4MIL-140, meeting both purity and recovery targets.

The traditional materials are comparable in purity to F4MIL-140 at low pressures due to the steep desorption gradients, however implementation of F4MIL-140 would greatly increase the working capacity and recovery while reducing the specific energy at these operating conditions, improving overall process productivity when compared to traditional materials.

7.3 PVSA

The PVSA cycle provides the strongest performance for all materials, satisfying purity and recovery requirements as the entire range of the isotherm is exploited.

Despite the comparable purity/recovery between traditional adsorbents and STAs, the specific energy and working capacity associated with implementing STA materials offer greater process advantages. For a given intermediate pressures, F4MIL-140 exhibits a far greater working capacity, while ZIF-7 provides a far lower specific energy than traditional adsorbents. A combination of the two materials in the same bed could reduce energy costs and increase working capacity even further, optimising the process.

However, despite these considerations, traditional materials are comparable and could potentially be practical for application. Although the STAs perform better, the lower cost of traditional materials could potentially make them competitive for the PVSA cycle. Zeolite 13X, for example, achieves 70.7% recovery and 95.4% purity, and although the recovery for this material is not as high as that of the STA's, Zeolite 13X is usually significantly cheaper than MOFs like F4MIL-140 with commercial Zeolite 13X being approximately \$0.5/kg and F4MIL-140 supposedly being 10-1000 times more expensive.²³

7.4 Process Considerations

When selecting between PSA, VSA and PVSA operation for STAs, a suitable material should be selected to ensure the phase change phenomena is fully exploited. Doing so results in greater KPI performance, especially reducing the specific energy and increasing the working capacity.

VSA operation imposes the largest energetic penalty as maintaining vacuum conditions requires significant work. For STAs, the additional CO₂ desorption achieved under vacuum does not sufficiently offset this energy cost, making VSA less attractive. Although VSA avoids the compression work associated with PVSA, the reduced amount of CO₂ recovered per cycle does not compensate for the high work associated with operating in vacuum conditions.

PSA demonstrates high purity and recovery are achievable without the need to operate in vacuum conditions. Additionally, when comparing the optimal STA's in both PSA and VSA conditions, ZIF-7 and F4MIL-140 respectively, it is important to note that energy consumption for the VSA cycle is almost 9 times larger than the PSA cycle, a factor that could give PSA a huge process advantage. Overall, the PSA cycle demonstrates that high process performance is now achievable with suitable material selection, with large CO₂ desorption occurring while having greater energy savings.

PVSA offers the greatest working capacity among all cycles as it utilises the full range of the isotherm & phase transition. This enhanced working capacity can reduce equipment sizing of the process and subsequent capital costs, making it a promising option despite the energy associated with vacuum operation.

Ultimately, the selection of operating conditions will depend on the downstream application, where higher CO₂ purity may be necessary for pipeline transport and geological sequestration, where non-condensable impurities (ie, N₂, O₂, H₂O and SO₂) can cause corrosion and reduce compression efficiency.²² Conversely, maximising recovery may be desirable where CO₂ emissions incur financial penalties, such as carbon taxes.

7.5 Processing time

Implementation of the STA equations while successful, has highlighted an issue associated with screening these materials, as each iteration of the process takes significantly longer to run than traditional isotherms, in some cases exceeding 2 minutes. This is primarily due to requiring small pressure increments when solving the BAAM ODE to accurately calculate the desorption path of the STA.

This limitation makes screening of many STAs challenging, as greater processing time and

computational power is required to compare materials, leading to greater costs²³.

Additionally, integrating STAs within more complex adsorption models would increase run times even further, making it more difficult to determine an optimum adsorbent for a given process.

8. Conclusion

Structural transition adsorbents are a promising class of materials for post-combustion carbon capture due to their distinct phase change behaviour, improving the key performance indicators of the cyclical carbon capture adsorption process. This gives them a unique advantage over their traditional counterparts. Amongst the case studies, PVSA demonstrated the greatest capacity to exploit the physical properties of the STA, offering high purity and recovery, greater working capacities and low specific energy. Additionally, it was found operating under vacuum conditions is not necessary to obtain high purity, recovery and working capacities, a key consideration in light of energy costs associated with pursuing a VSA cycle.

While the results of this report favour STA materials, several challenges must be addressed before a definitive material selection can be made. Adsorption models such as BAAM fail to capture the often-exothermic nature associated with the phase change of STA materials, resulting in non-isothermal operation and affecting the location and size of the step on the equilibrium isotherm. Furthermore, increased computational demand and processing time associated with numerical integration of STAs into process models limit the ability to screen large databases of materials.

Future work should focus on expanding material screening, where STAs with multiple phase changes may provide greater process performance, alongside incorporating more accurate process models to assess material performance. Through such efforts, the full potential of STA materials can be realised, and a definitive assessment of their suitability can be conducted.

Acknowledgements

The authors of this report would like to express immense thanks to Dr. Hassan Azzan and Ms. Kristen He for their support and guidance throughout this period.

References

- [1] US Department of Commerce, N. (n.d.). *Global Monitoring Laboratory - Carbon Cycle Greenhouse Gases*. Gml.noaa.gov. <https://gml.noaa.gov/ccgg/trends/mlo.html>
- [2] Muhammad Kabir, Um E Habiba, Wali Khan, Amin Shah, Sarvat Rahim, Patricio R. De los Rios-

Escalante, Zia-Ur-Rehman Farooqi, Liaqat Ali, Muhammad Shafiq, Climate change due to increasing concentration of carbon dioxide and its impacts on environment in 21st century; a mini review, *Journal of King Saud University - Science*, Volume 35, Issue 5, 2023, 102693, ISSN 1018-3647, <https://doi.org/10.1016/j.jksus.2023.102693>. (<https://www.sciencedirect.com/science/article/pii/S1018364723001556>)

[3] - Energy Institute. (2025). *Statistical review of world energy 2025* (74th ed.). Energy Institute.

[4] COWI. (2024). *Comparing CO₂ emissions from different energy sources*. COWI. <https://www.cowi.com/news-andpress/news/2023/comparing-co2-emissions-from-different-energy-sources/>

[5] UNFCCC. (2015). *The Paris Agreement*. UNFCCC. https://unfccc.int/sites/default/files/resource/parisagreement_publication.pdf

[6] Reynolds, A. J., Verheyen, T. V., Adeloju, S. B., Meuleman, E., & Feron, P. (2012). Towards Commercial Scale Postcombustion Capture of CO₂ with Monoethanolamine Solvent: Key Considerations for Solvent Management and Environmental Impacts. *Environmental Science & Technology*, 46(7), 3643–3654. <https://doi.org/10.1021/es204051s>

[7] World Nuclear Association. (2021, November 16). *“Clean Coal” Technologies, Carbon Capture & Sequestration - World Nuclear Association*. World-Nuclear.org. <https://world-nuclear.org/information-library/energy-and-the-environment/clean-coal-technologies>

[8] Sharon Sjostrom, Constance Senior, Pilot testing of CO₂ capture from a coal-fired power plant—Part 1: Sorbent characterization, *Clean Energy*, Volume 3, Issue 2, June 2019, Pages 144–162, <https://doi.org/10.1093/ce/zkz009>Top of FormBottom of Form

[9] Li, J.-R., Kuppler, R. J., & Zhou, H.-C. (2009). Selective gas adsorption and separation in metal–organic frameworks. *Chemical Society Reviews*, 38(5), 1477. <https://doi.org/10.1039/b802426j>

[10] Adam Ward, Ke Li, Ronny Pini, Assessment of dual-adsorbent beds for CO₂ capture by equilibrium-based process design, *Separation and Purification Technology*, Volume

- 319, 2023, 123990, ISSN 1383-5866,
<https://doi.org/10.1016/j.seppur.2023.123990>.
 (https://www.sciencedirect.com/science/article/pii/S1383586623008985)
- [11] Senkovska, I., Bon, V., Mosberger, A., Wang, Y., & Kaskel, S. (2025). Adsorption and Separation by Flexible MOFs. *Advanced Materials*.
<https://doi.org/10.1002/adma.202414724>
- [12] S. Hiraide, Y. Sakanaka, Y. Iida, H. Arima, M.T. Miyahara, & S. Watanabe, Theoretical isotherm equation for adsorption-induced structural transition on flexible metal–organic frameworks, *Proc. Natl. Acad. Sci. U.S.A.* 120 (31) e2305573120,
<https://doi.org/10.1073/pnas.2305573120> (2023).
- [13] Wang, J., Wei, Y., & Ma, Z. (2020). Modified Dual-Site Langmuir Adsorption Equilibrium Models from A GCMC Molecular Simulation. *Applied Sciences*, 10(4), 1311–1311.
<https://doi.org/10.3390/app10041311>
- [14] Vishal Subramanian Balashankar, Rajagopalan, A. K., Pauw, R. de, Avila, A. M., & Rajendran, A. (2019). Analysis of a Batch Adsorber Analogue for Rapid Screening of Adsorbents for Postcombustion CO₂ Capture. *Industrial & Engineering Chemistry Research*, 58(8), 3314–3328.
<https://doi.org/10.1021/acs.iecr.8b05420>
- [15] 9.2. Carbon Dioxide Capture Approaches. (n.d.). Netl.doe.gov.
<https://netl.doe.gov/research/carbon-management/energy-systems/gasification/gasifiedia/capture-approaches>
- [16] Danaci, D., Pulidori, E., Bernazzani, L., Petit, C., & Taddei, M. (2023). Evaluating the CO₂ capture performance of a “phase-change” metal–organic framework in a pressure-vacuum swing adsorption process. *Molecular Systems Design & Engineering*, 8(12), 1526–1539.
<https://doi.org/10.1039/d3me00098b>
- [17] – Klein, R. A., Shulda, S., Parilla, P. A., Le Magueres, P., Richardson, R. K., Morris, W., Brown, C. M., & McGuirk, C. M. (2021). Structural resolution and mechanistic insight into hydrogen adsorption in flexible ZIF-7 [Supplementary information]. *Chemical Science*, 12(47), 15620–15631.
<https://doi.org/10.1039/D1SC04618G>
- [18] Lanlan Zhang, Kai Han, Yongzhen Wang, Baojun Tang, Junyao Wang, Qike Han, Jiayu Lin, A comparative experimental study on energy and exergy performance of cascade electrical-temperature vacuum swing adsorption (C-ETVSA) for carbon capture, *Chemical Engineering Journal*, Volume 521, 2025, 166593, ISSN 1385-8947,
<https://doi.org/10.1016/j.cej.2025.166593>.
 (https://www.sciencedirect.com/science/article/pii/S1385894725074315)
- [19] – Leung, D. Y. C., Caramanna, G., & Maroto-Valer, M. M. (2014). An overview of current status of carbon dioxide capture and storage technologies. *Renewable and Sustainable Energy Reviews*, 39, 426–443.
<https://doi.org/10.1016/j.rser.2014.07.093>
- [20] *Point Source Carbon Capture from Industrial Sources*. (n.d.). Netl.doe.gov.
<https://netl.doe.gov/carbon-capture/industrial>
- [21] Bacilieri, A., Black, R., & Way, R. (2023). *Assessing the relative costs of high-CCS and low-CCS pathways to 1.5 degrees*.
<https://www.smithschool.ox.ac.uk/sites/default/files/2023-12/Assessing-the-relative-costs-of-high-CCS-and-low-CCS-pathways-to-1-5-degrees.pdf>
- [22] *Guidance on conveying carbon dioxide in pipelines in connection with carbon capture and storage projects - HSE*. (2025). Hse.gov.uk.
<https://www.hse.gov.uk/pipelines/co2conveying-full.htm>
- [23] Farmahini, A. H., Krishnamurthy, S., Friedrich, D., Brandani, S., & Sarkisov, L. (2021). Performance-Based Screening of Porous Materials for Carbon Capture. *Chemical Reviews*, 121(17), 10666–10741.
<https://doi.org/10.1021/acs.chemrev.0c01266>

k-BAAM: A kinetics-augmented batch adsorber analogue model for sorbent screening in pressure-vacuum swing adsorption process for point-source CO₂ capture*

Alice Zhu*, Eric Lun*

*Department of Chemical Engineering, Imperial College London, London, SW7 2AZ, UK

Global reliance on fossil fuels makes carbon dioxide (CO₂) capture essential for climate-change mitigation, yet adsorption-based pressure vacuum swing adsorption (PVSA) process models either neglect adsorption kinetics or are too computationally expensive. This work implements and deploys a kinetics-augmented Batch Adsorber Analogue Model (k-BAAM) that includes time derivatives and adsorption kinetics, which relax the instantaneous equilibrium assumption and incorporates a Linear-Driving-Force-based mass-transfer term. Hence, the k-BAAM model delivers physically reliable estimates of CO₂ purity, recovery, productivity and energy consumption at ~60x faster speed than a high-fidelity model. Seven adsorbents – Zeolite 13X, UTSA-16, CALF-20, MIL-160(Al), CuBTC, Activated Carbon and Mg-MOF-74 – are examined with UTSA-16 identified as the top performer. While equilibrium descriptors, including isotherm capacity, working capacity and selectivity, govern separation, adsorption kinetics in the form of mass-transfer efficiency governs productivity and energy use. A top-performing adsorbent must therefore combine favourable equilibrium thermodynamics with rapid adsorption kinetics.

Key Words: k-BAAM, adsorbent screening, adsorption kinetics, KPI, adsorbents, PVSA, CO₂ capture

Nomenclature

c_i	Gas-phase concentration of species i
C_{eff}	Effective heat capacity (HiFi)
D_{ax}	Axial dispersion coefficient (HiFi)
En	Specific energy consumption
k	LDF mass-transfer coefficient
M_{CO_2}	Molar mass of CO ₂
n_i	Total moles of species i in the system
p_H	Adsorption pressure
p_I	Blowdown pressure
p_L	Evacuation pressure
Pr	Productivity
Pu	CO ₂ purity
q	Solid-phase loading
q^*	Equilibrium solid-phase loading
Q_{ads}	Heat of adsorption
Q_{loss}	Heat-loss term
Re	CO ₂ recovery
t_{step}	Duration of a specific cycle step
t_{cycle}	Total cycle time
T	System temperature
u	Gas velocity (HiFi)
W	Instantaneous specific work
X_1	Parameters for CO ₂
X_2	Parameters for N ₂
y_i	Gas-phase mole fraction of species i
γ	Heat-capacity ratio
η	Efficiency factor (isentropic/mechanical)
λ_{eff}	Effective thermal conductivity (HiFi)

1. Introduction

Global reliance on fossil fuels has made carbon dioxide (CO₂) the dominant contributor to global warming.¹⁾ For as long as the global energy system is dependent on fossil fuels, its capture will remain a central element in climate-change mitigation. To achieve decarbonisation goals,²⁾ power plants and industrial facilities require technologies capable of selectively removing CO₂ either from flue gas or directly from ambient air. Gas-phase separation by adsorption onto porous solid sorbents is an attractive route to meet this need while offering high selectivity,³⁾ straightforward implementation in processes,⁴⁾ and economic advantages.⁵⁾

Industrial adsorption processes are commonly operated as pressure- (P-), vacuum- (V-), or temperature- (T-) swing adsorption (-SA) cycles, in which the bed pressure and/or temperature is modulated to promote CO₂ uptake and desorption.⁶⁾ For the simulation and optimisation of such processes, several high-fidelity (HiFi) dynamic process models have been developed that solve mass- and energy-balance partial differential equations (PDEs) before being embedded within black-box optimisation frameworks.⁷⁾ Although the models provide high accuracy, optimising a single process configuration under a large number of decision variables can require several days of computational effort⁸⁾ with each simulation taking up to 5 minutes to converge.⁹⁾ This computational cost becomes a limiting factor when the aim becomes screening multiple adsorbent candidates.

To lower computational time, simplified shortcut models

have been proposed. A prominent example is the batch adsorber analogue model (BAAM), a zero-dimensional (0-D) equilibrium-based proxy that assumes perfect mixing and instantaneous solid–fluid equilibrium.¹⁰⁾ The BAAM model enables rapid estimation of process-level performance, achieving less than 1 second for each data point. Based solely on equilibrium isotherm data, it has proved useful for preliminary material and process screening.¹¹⁾ However, such equilibrium formulations neglect adsorption kinetics, as modern sorbents – particularly metal–organic frameworks (MOFs) and other advanced porous solids – often exhibit diffusion resistances where the bed may not achieve equilibrium within the timescale of typical cycle steps.¹⁰⁾

To address the limitations, a kinetics-enabled BAAM model (k-BAAM) is developed. The model builds on the previous BAAM framework, incorporates time derivatives, and accounts for adsorption kinetics by introducing a linear driving force (LDF). Time derivatives consider the time taken for each cycle to reach cyclic steady state, relaxing the instantaneous equilibrium assumption. The LDF expression describes adsorption kinetics, parameterises material-specific equilibrium loadings, and incorporates effective mass-transfer coefficients. The model is then formulated for a four-step pressure-vacuum swing adsorption (PVSA) cycle and models point-source CO₂ capture in which the CO₂ concentration varies from a CO₂/N₂ binary mixture.

This report presents the implementation and usage of the k-BAAM model for adsorbent screening at point-source CO₂ concentrations of 5%, 10%, 15%, and 20%. Key Performance Indicators (KPIs) of CO₂ purity, CO₂ recovery, productivity, and specific energy consumption are integrated so that the benefits and trade-offs across different adsorbents can be quantified, analysed, and compared. In this work, seven promising adsorbents - Zeolite 13X, CALF-20, Mg-MOF-74, Activated Carbon, Cu-BTC, UTSA-16, and MIL-160 - are examined.

2. Methodology

This section details the PVSA cycle, the construction of the k-BAAM, the definitions of KPIs and objective functions, leading up to the optimisation framework and selection of optimal operating points.

2.1. PVSA Process Cycle Description

A four-step PVSA cycle for post-combustion CO₂ capture, as shown in Figure 1, is considered for the k-BAAM. The cycle consists of:

- (1) Adsorption step: Flue gas (a binary mixture of CO₂/N₂) is fed into the adsorption column at p_H . CO₂ is selectively adsorbed onto the solid adsorbent while N₂-rich gas exits from the product end.
- (2) Blowdown step: The inlet is closed, and the column pressure is lowered to an intermediate level p_I by removing a portion of the gas remaining in the column.
- (3) Evacuation (Desorption) step: The column is further

depressurised to a vacuum pressure p_L for adsorbed CO₂ to be released.

- (4) Repressurisation step: The column is brought back to p_H by directing the N₂ back to the chamber.

Steps (1)–(4) are then repeated in cycles until the process reaches CSS. In simulations, CSS is achieved when the changes in key state variables between consecutive cycles fall below a tolerance of 10^{-3} .

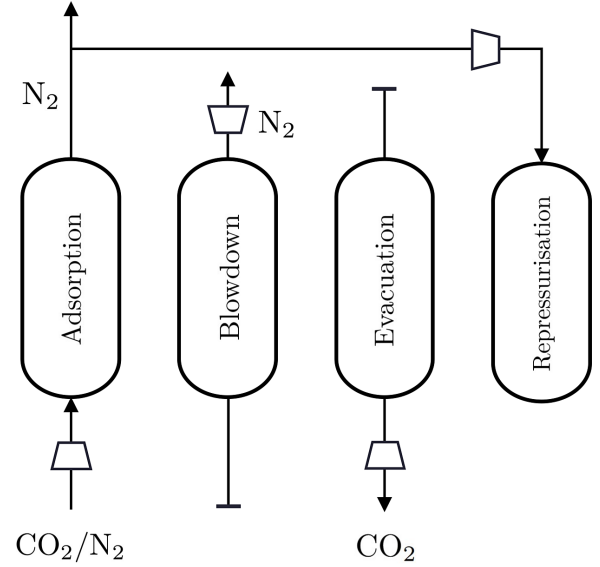


Fig. 1. Schematic for four-step light product pressurisation PVSA cycle for point-source CO₂ capture.

2.2. Adsorbent Materials and Properties

The following set of seven adsorbent materials is examined in the k-BAAM model for the following reasons:

- **Zeolite 13X** : a benchmark molecular sieve for CO₂ capture from flue gas.¹²⁾
- **Activated Carbon**: a widely available commercial adsorbent with hydrophobic properties.
- **Mg-MOF-74** : a promising MOF material for CO₂ capture; can be used in combination with Activated Carbon.¹³⁾
- **UTSA-16** : a top performing diamine-functionalised MOF for post-combustion CO₂ capture.¹⁴⁾
- **CuBTC** (or HKUST-1): a promising MOF and the first MOF produced at an industrial scale by BASF.
- **MIL-160**: a microporous aluminium fumarate MOF, one of the few MOFs tested at pilot-scale for post-combustion CO₂ capture.
- **CALF-20**: the only MOF produced at commercial scale for post-combustion CO₂ capture.¹⁵⁾

The selected materials cover a range of adsorbents with varied combinations of equilibrium CO₂ capacities, selectivities and transport properties. This enables an initial assessment of equilibrium-kinetic trade-offs. The input data of their parameter adsorbents can be found in Table 1, where they are held constant across the 7 adsorbents.

Table 1. Material properties and process operating point.

Symbol	Description	Value [unit]
$c_{p,g}$	Gas heat capacity	30.7 J mol ⁻¹ K ⁻¹
$c_{p,s}$	Solid heat capacity	1070 J kg ⁻¹ K ⁻¹
F_{in}	Inlet molar flow	0.04036 mol s ⁻¹
h_{wall}	Wall-fluid heat-trans. coeff.	10 W m ⁻² K ⁻¹
r_{in}	Column inner radius	0.05 m
r_p	Pellet radius	0.0015 m
T_{amb}	Ambient temperature	298 K
T_{feed}	Feed temperature	298 K
V	Column/bed volume	0.01 m ³
$y_{CO_2,in}$	Feed CO ₂ mole fraction	0.15
ε_b	Bed void fraction	0.38
ρ_s	Solid density	1130 kg m ⁻³

2.3. Material descriptors

Two categories of material descriptors are incorporated: equilibrium and kinetic. The equilibrium descriptors include isotherm uptake q_1^* , working capacity Δq_{CO_2} , and selectivity $S_{ads,CO_2/N_2}$; the kinetic descriptor is mass transfer coefficient k_1 .

2.3.1. Equilibrium descriptors

Adsorption equilibrium is described by an extended Dual Site Langmuir (DSL) model. The equilibrium CO₂ loading $q_1^*(c_1, c_2)$ at T is:

$$q_1^*(c_1, c_2) = q_{sb,1} \frac{b_1(T)c_1}{1 + b_1(T)c_1 + b_2(T)c_2} + q_{sd,1} \frac{d_1(T)c_1}{1 + d_1(T)c_1 + d_2(T)c_2} \quad (1)$$

$$c_i = \frac{Py_i}{RT}, i = 1, 2. \quad (2)$$

$$b_i(T) = b_{0,i} \exp\left(-\frac{\Delta U_{b,i}}{RT}\right) \quad (3)$$

$$d_i(T) = d_{0,i} \exp\left(-\frac{\Delta U_{d,i}}{RT}\right) \quad (4)$$

Selectivity and working capacity are obtained from:

$$S = \frac{q_1^*/q_2^*}{y_1/y_2} \quad \text{and} \quad \Delta q_1 = q_{1,ads}^* - q_{1,des}^* \quad (5)$$

2.3.2. Kinetic descriptor

Mass transfer is modelled by an LDF expression derived from macropore diffusion in spherical pellets. An effective LDF coefficient, or mass transfer coefficient k_1 , for

Table 2. Process parameters and decision variables

Parameter	Unit	BAAM	k-BAAM/HiFi
Decision Variables			
Intermediate P_I	bar	[0.2, 0.8]	[0.2, 0.8]
t_{ads}, t_{evac}	s	–	[20, 1500]/[60, 240]
Fixed Input Parameters			
High Press. (P_H)	bar	1.00	1.00/1.01
Low Press. (P_L)	bar	0.01	0.01

CO₂ is defined as below, accounting for both diffusional resistance and local storage capacity.

$$k_1(c_1, c_2) = K_0 \frac{c_1/\rho_s}{q_1^*(c_1, c_2)}, \quad \text{where } K_0 = \frac{15\varepsilon_p D_m}{\tau_p r_p^2} \quad (6)$$

2.4. Process parameters and decision variables.

Table 2 shows the process parameters and decision variables chosen for k-BAAM. Four CO₂ feed concentrations are chosen: $y_{CO_2,feed} \in \{0.05, 0.10, 0.15, 0.20\}$, p_H, p_L were taken as fixed input parameters, and p_I, t_{ads}, t_{evac} as decision variables with upper and lower boundaries. The cycle time is

$$t_{cycle} = t_{ads} + t_{bd} + t_{evac} + t_{press} \quad (7)$$

2.5. Model comparisons: BAAM, k-BAAM, and High-Fidelity Model

BAAM and k-BAAM are both zero-dimensional (0-D) equilibrium models assuming no spatial concentration or temperature gradients along the column. The entire bed in the model is treated as a single well-mixed unit, without axial variation in gas or solid phase properties.¹¹⁾ k-BAAM, in addition, incorporates adsorption kinetics and time derivatives. For each step, k-BAAM solves coupled ordinary differential equations (ODE) for the moles of CO₂ and N₂ in the gas phase and their adsorbed amounts, using a LDF rate law based on the material-specific equilibrium loading q_1^* and effective kinetic coefficient k_1 . Integration proceeds over repeated cycles until CSS is reached. Given that BAAM and k-BAAM share identical equilibrium, vacuum and operating-pressure descriptions, differences in performance can be attributed to kinetic limitations. For the HiFi Model,¹⁶⁾ the gradients and axial variation are considered, hence a 1-D model. The spatial and temporal profiles in the column are obtained by solving PDEs for mass, energy and momentum balances. Table 3 summarises the key equations for each model.

2.6. Performance indicators

Four KPIs are considered in this report: purity (Pu), recovery (Re), specific energy (En), and productivity (Pr).

2.6.1. Purity

CO₂ purity is defined as

$$Pu = 100 \frac{n_{CO_2,prod}}{n_{CO_2,prod} + n_{N_2,prod}} [\%]. \quad (8)$$

where $n_{i,prod} = n_{i,bd} - n_{i,evac}$.

2.6.2. Recovery

CO₂ recovery is defined as

$$Re = 100 \frac{n_{CO_2,prod}}{n_{CO_2,in}} [\%]. \quad (9)$$

where $n_{CO_2,in} = F_{in} y_{CO_2,feed} t_{ads}$.

2.6.3. Productivity

The dynamic volumetric productivity is defined as

$$Pr = \frac{n_{CO_2,prod}}{V_{bed} t_{cycle}} [\text{mol m}^{-3} \text{s}^{-1}], \quad (10)$$

Table 3. Model formulation summary of BAAM, k-BAAM and HiFi.

Model	Equation Name	Complete Equation
BAAM (0D isothermal)	Equilibrium sorption	$q_i = q_i^*(P, y, T)$
	Material balance	$Total\ in = Total\ mass\ out, \quad CO_2\ in = CO_2\ out$
	Blowdown mapping	$\frac{dy}{dP} = f(P, y)$
k-BAAM (0D isothermal)	Solid-phase kinetics (LDF)	$\frac{dq_i}{dt} = k_i(P(t), y(t)) (q_i^*(P(t), y(t), T) - q_i(t))$
	Gas-phase balance	$\varepsilon V \frac{d}{dt} \left(\frac{P(t) y(t)}{RT} \right) = F_{in}(t) y_{in} - F_{out}(t) y(t) - (1 - \varepsilon) V \frac{dq_{CO_2}}{dt}$
HiFi (1D non-isothermal)	Solid-phase kinetics (LDF)	$\frac{\partial q_i}{\partial t}(z, t) = k_i(z, t) (q_i^*(P(z, t), y(z, t), T(z, t)) - q_i(z, t))$
	Gas-phase species PDE	$\varepsilon \frac{\partial C_i}{\partial t} + u(z, t) \frac{\partial C_i}{\partial z} = \varepsilon D_{ax} \frac{\partial^2 C_i}{\partial z^2} + (1 - \varepsilon) \rho_p \frac{\partial q_i}{\partial t}$
	Energy PDE	$C_{eff} \frac{\partial T}{\partial t} + u(z, t) \frac{\partial T}{\partial z} = \lambda_{eff} \frac{\partial^2 T}{\partial z^2} + Q_{ads}(z, t) - Q_{loss}(z, t)$
	Pressure drop	$\frac{\partial P}{\partial z} = -(A u(z, t) + B u(z, t) u(z, t))$

2.6.4. Specific energy consumption

Specific energy is the electrical energy for gas handling per tonne of CO₂ recovered:

$$En = \frac{W_{vac} + W_{comp}}{m_{CO_2, prod}}, \quad (11)$$

where $m_{CO_2, prod} = n_{CO_2, prod} M_{CO_2}$, and $W_{comp} = 0$. The vacuum model assumes ideal-gas, adiabatic compression with $\gamma = 1.4$, $T = 298$ K and $P_{dis} = 1$ bar. The compressor model is consistent with other point-source PVSA work.¹⁷⁾ The isentropic efficiency at $P_{suc}(t)$ is:

$$\eta_{vac}(P_{suc}) = \eta_{max} \frac{a(\frac{P_{suc}}{10^5})}{1 + a(\frac{P_{suc}}{10^5})}, \quad (12)$$

where $a = 19.55$, $\eta_{max} = 0.80$. The instantaneous specific work:

$$w_{vac}(t) = \frac{\gamma}{\gamma - 1} RT \eta_{vac}(P_{suc}(t)) \left[\left(\frac{P_{dis}}{P_{suc}(t)} \right)^{\frac{\gamma-1}{\gamma}} - 1 \right] \quad (13)$$

For step j (blowdown or evacuation) of duration t_j , with molar outflow $F_{out,j}(t)$, the vacuum energy is

$$W_j = \frac{1}{3.6 \times 10^6} \int_0^{t_j} F_{out,j}(t) w_{vac}(t) dt, \quad (14)$$

Total vacuum energy per cycle is $W_{vac} = W_{bd} + W_{evac}$.

2.7. Objective Functions

Two multi-objective optimisation cases are solved: Separation case, Pu-Re domain

$$\min_{x \in X} f_1(x) = (-Pu(x), -Re(x)). \quad (15)$$

Cost case, En-Pr domain

$$\min_{x \in X} f_2(x) = (En(x), -Pr(x)) \quad s.t. \quad Pu(x) \geq 80\%. \quad (16)$$

In both multi-objective optimisation cases, the analysis focuses exclusively on the Pareto fronts (PF). A PF constitutes a set of non-dominated solutions. A point is Pareto-optimal if no other feasible design improves one KPI (e.g. higher Pu, higher Re, lower En, or higher Pr) without degrading at least one other.

2.8. Global optimisation: Sobol sampling and Non-dominated Sorting Genetic Algorithm II (NSGA-II)

Sobol Sampling generates operating points that uniformly fill the defined multi-dimensional variable domain to provide a global, unbiased exploration of the knowledge space. In this work, a MATLAB configuration of *rng(7, 'twister')*; *Skip=1000*, *Leap73* is used. A multi-objective genetic algorithm, NSGA-II, evolves a population (pop) of solutions towards the PF via numerous generations (gen) by keeping non-dominated points. BAAM uses only Sobol sampling; HiFi Model uses only NSGA-II; k-BAAM, on the other hand, incorporates both to obtain its PFs. For k-BAAM, NSGA-II uses the feasible points from Sobol as the seed to refine the PFs. Table 4 details the optimisation configurations for the 3 models.

Table 4. Optimisation configurations.

Model	Sobol samples	NSGA-II
BAAM	2048	-
k-BAAM	8192	Pop=140, Gen=100
HiFi	-	Pop=72, Gen=50

2.9. Definition of the optimal point

For clear comparison across adsorbents, the best-performing point at (Pu^*, Re^*) is selected from the Pu-Re domain. The optimal point is defined as having the shortest straight-line distance from $(Pu, Re) = (100, 100)$:

$$\min_{f \in F_1(x)} D = \sqrt{(Pu - 100)^2 + (Re - 100)^2} \quad (17)$$

3. Results

This section discusses the credibility of the k-BAAM model, the implications of feed concentration, the performance of adsorbents alongside the impact of material descriptors, and the significance of the k-BAAM model.

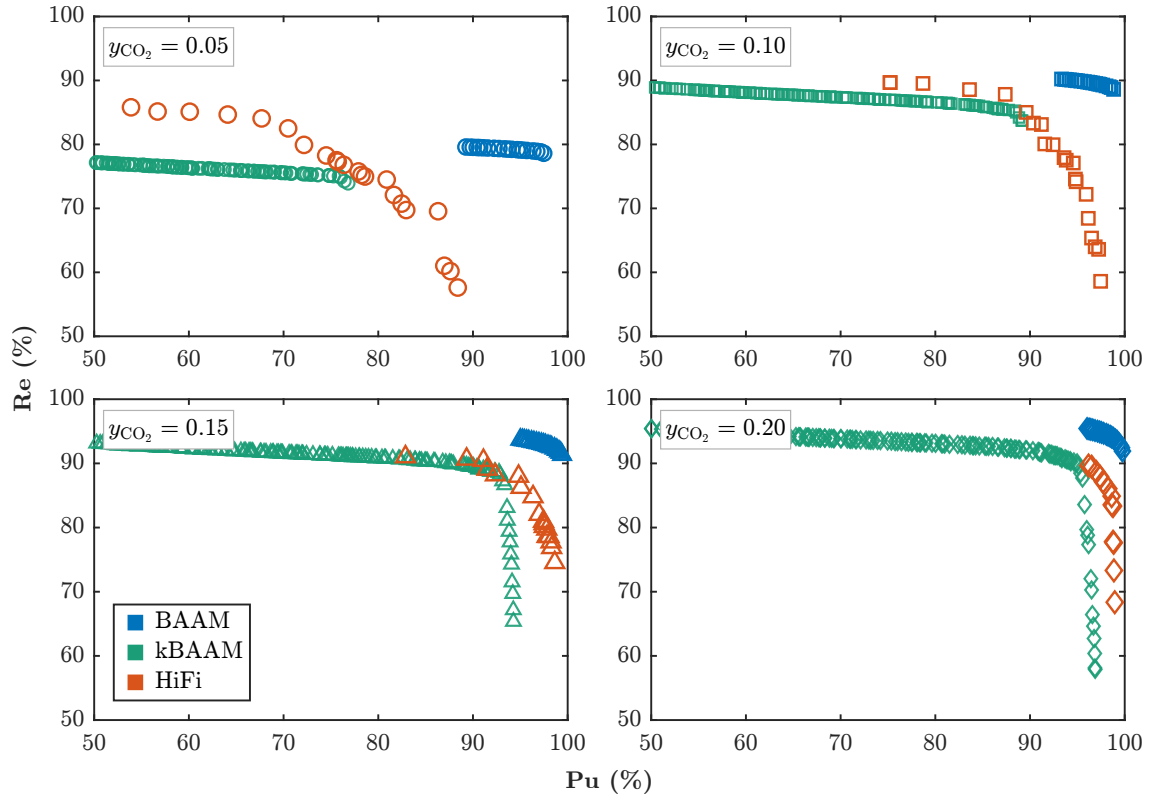


Fig. 2. Model comparisons: Pareto fronts for BAAM, k-BAAM, and HiFi Models across different feed concentrations.

3.1. Model comparisons

As shown in Figure 2, k-BAAM was first compared on the Pu-Re domain against the PFs obtained from BAAM and the HiFi Model for Zeolite 13X across the 4 inlet CO_2 feeds.

As shown, BAAM systematically overestimates process performance as it enforces instantaneous solid-gas equilibrium and neglects mass-transfer kinetics. The results produced by BAAM represent an idealised upper performance limit given its assumption of the sorbents constantly operating at equilibrium capacity.

In contrast, k-BAAM relaxed the equilibrium constraint and incorporated adsorbent kinetics, closely reproducing the curvature and extent of the HiFi Pu-Re PFs with small, non-systematic deviations and slightly less complete PFs.

The HiFi Model has been shown to reproduce KPI

trends in close agreement with literature simulations,¹⁶⁾ hence describing realistic performance. It is also noted that while a singular data point requires ~ 5 min of computation time for the HiFi Model, it takes 5sec on the k-BAAM model, achieving 60x faster than the HiFi Model and reducing over 98% of computational time.

The combination of reasonable accuracy, non-exaggerated performance, and low computational cost makes k-BAAM a credible surrogate for the HiFi Model and an appropriate tool for adsorbent screening and design-space exploration.

3.2. Equilibrium and Kinetic Descriptors

Figure 3(a) and (b) show the behaviour of q^* and k_1 , respectively, for all 7 adsorbents across CO_2 feed concentrations from 0 to 30%. As observed, with the increase of y_{CO_2} , all materials show an increase in both q^* and k_1 , but with different extents and curvature. From

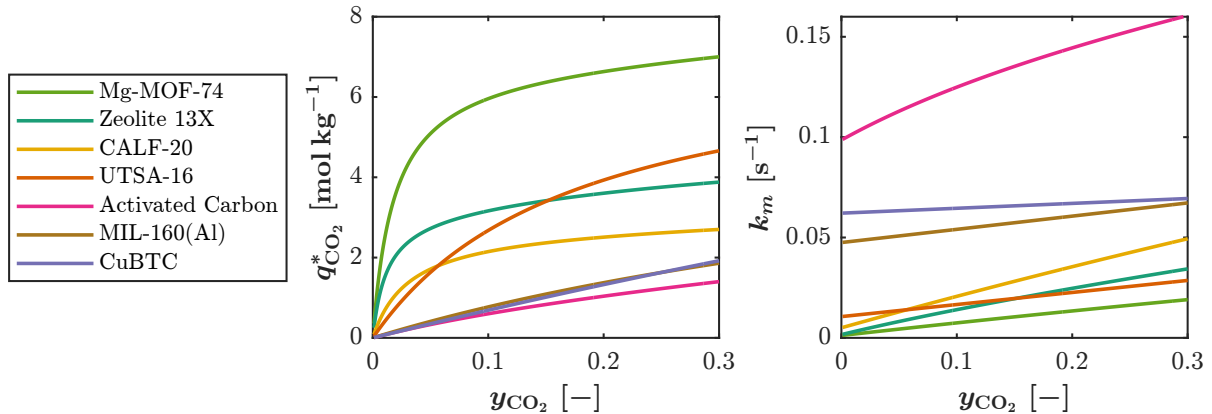


Fig. 3. Equilibrium and kinetic descriptors for adsorbents.

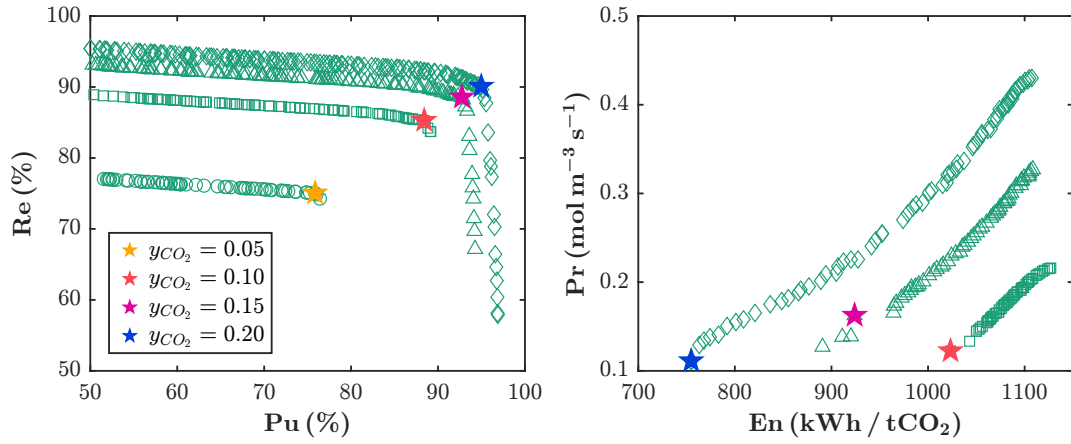


Fig. 4. Effect of feed composition on Zeolite 13X performance. Starred points represent the best performing points defined in 2.9.

Figure 3(a), it can be noted that materials including Mg-MOF-74, Zeolite 13X, CALF-20, and UTSA-16 depict steeply increasing and highly curved isotherms at low y_{CO_2} , hence achieving high loadings. These indicate strong CO_2 affinity and large saturation capacities. The other materials, Activated Carbon, MIL-160(Al), and CuBTC, exhibit gradual, almost linear behaviour over the same range, reflecting weaker affinity to and lower capacities for CO_2 . The equilibrium uptake for each adsorbent is inversely proportional to its effective kinetic coefficient, as shown before via Eqn (6):

$$k_m(c_1, c_2) = K_0 \frac{c_1/\rho_s}{q_1^*(c_1, c_2)}$$

Hence, materials with larger equilibrium capacities necessarily return smaller k_1 , while low-capacity adsorbents, such as Activated Carbon, produce high k_1 . This is an intrinsic consequence of the LDF: for a fixed concentration driving force, a sorbent that stores more CO_2 per pellet volume appears kinetically slower, whereas one with smaller storage appears faster.

3.3. Zeolite 13X: effect of feed composition

Figure 4(a) and (b) shows the PFs of Zeolite 13X on the Pu-Re domain and the En-Pr domain, respectively. The coloured stars represent the optimal points. In the Pu-Re domain across different inlet feed concentrations, Re slowly decreases with the increase of Pu from $\sim 50\%$ to $\sim 90\%$. This indicates that Pu can increase significantly with a minute penalty on Re. However, once Pu for $y_{\text{CO}_2} = 0.15, 0.20$ pushes beyond $\sim 90\%$, Re drops sharply. It can be inferred that achieving extreme Pu requires disproportionately large sacrifices in Re. The En-Pr PFs, on the other hand, show almost a linear trend and a clear trade-off, suggesting an increase in Pr would lead to a proportional increase in En. Across the four inlet feed concentrations examined, as the feed becomes richer in CO_2 , the Pu-Re PFs shift monotonically towards the upper-right corner. It is also observed that at feeds of $y=15\%$ and 20% , the Pu-Re frontier becomes more complete.

Under an 80% Pu constraint for Pu-Re, the PF for $y = 0.05$ is not depicted as no viable data from the concentration exhibits $\text{Pu} \geq 80\%$. It is also observed that

with the increase of CO_2 concentration in the feed, the En-Pr front moves monotonically towards the upper-left corner.

The optimal points selected from and starred on the Pu-Re domain are also highlighted on the En-Pr domain, where, incidentally, they land close to the $En_{\min}$. This indicates that within the decision space, designs that drive Pu and Re high must sacrifice Pr to avoid excessively long step times that lead to higher $En_{\min}$.

Note that Pu, Re, Pr values are of similar magnitudes to other PVSA studies. However, a pessimistic En is seen in this research: other simulations achieved En of $100-250 \text{ kWh/t}$,^{17,18)} significantly lesser than the $\sim 1000 \text{ kWh/t}$ observed in here due to the choice of a p_L of 0.01 bar .

The same simulations and analysis were conducted on the remaining six adsorbents. The same conclusions are arrived at.

3.4. Screening results: KPIs at $y=0.15$

Table 5 summarises the key data at $y_{\text{CO}_2, \text{feed}} = 0.15$ for the seven adsorbents. The adsorbents are listed in the order of increasing k_1 . The following are reported for each material: Equilibrium and kinetic descriptors q^* , Δq_{CO_2} , $S_{\text{ads}, \text{CO}_2/\text{N}_2}$ and k_1 , Pu^* , Re^* : the Pu and Re values obtained at the best performing point on the Pu-Re front, t_{ads} , t_{evac} , p_I : values of input variables at optimal point, $Pr_{\max}$: maximum Pr satisfying $Pu \geq 80\%$, $En_{\min}$: minimum En satisfying $Pu \geq 80\%$.

3.4.1. Individual results

It is observed that, with the increase of k_1 , Δq_{CO_2} decreases. The p_I across the adsorbents at the best performing point are all 0.20 bar with varied adsorption and evacuation times, some significantly longer than others. For adsorbents with valid $Pr_{\max}$ and $En_{\min}$, the two KPIs show no significant correlation with the increase of k_1 besides the apparent outlier of CALF-20.

UTSA-16 and Zeolite 13X yield one of the highest Pu^* , Re^* and $Pr_{\max}$ given their combination of sufficient capacity, high selectivity, and kinetics. Though CALF-20 demonstrates adequate kinetics, its moderate Δq_{CO_2} and $S_{\text{ads}, \text{CO}_2/\text{N}_2}$ require deeper vacuum to suppress N_2 slip, resulting in the highest $En_{\min}$. Despite modest equilibrium strength, MIL-160(Al) produces high $Pr_{\max}$ and low $En_{\min}$ given its moderate k_1 enables short cycles

Table 5. Screening results at $y = 0.15$

Adsorbent	k_m [s ⁻¹]	Δq_{CO_2} [mol kg ⁻¹]	$S_{ads, CO_2/N_2}$ [-]	Pu* [%]	Re* [%]	Point [t_{ads}, t_{evac}, P_r]	Pr_{max} [mol/m ³ s]	En_{min} [kWh/t]
Mg-MOF-74	0.010	4.054	55.4	77.2	89.8	[756 s, 906 s, 0.202 bar]	—	—
UTSA-16	0.020	3.037	280.2	94.0	89.7	[476 s, 1389 s, 0.202 bar]	0.387	803.5
Zeolite 13X	0.020	1.970	176.0	92.8	88.6	[719 s, 1405 s, 0.201 bar]	0.327	890.7
CALF-20	0.028	1.728	80.0	87.4	79.9	[508 s, 1476 s, 0.207 bar]	0.207	1189.4
MIL-160(Al)	0.057	1.009	38.7	86.2	80.0	[356 s, 1022 s, 0.201 bar]	0.345	756.1
CuBTC	0.066	0.948	25.8	83.2	76.9	[356 s, 1022 s, 0.201 bar]	0.332	771.4
Activated Carbon	0.135	0.741	29.3	80.3	73.6	[230 s, 1192 s, 0.201 bar]	0.278	726.3

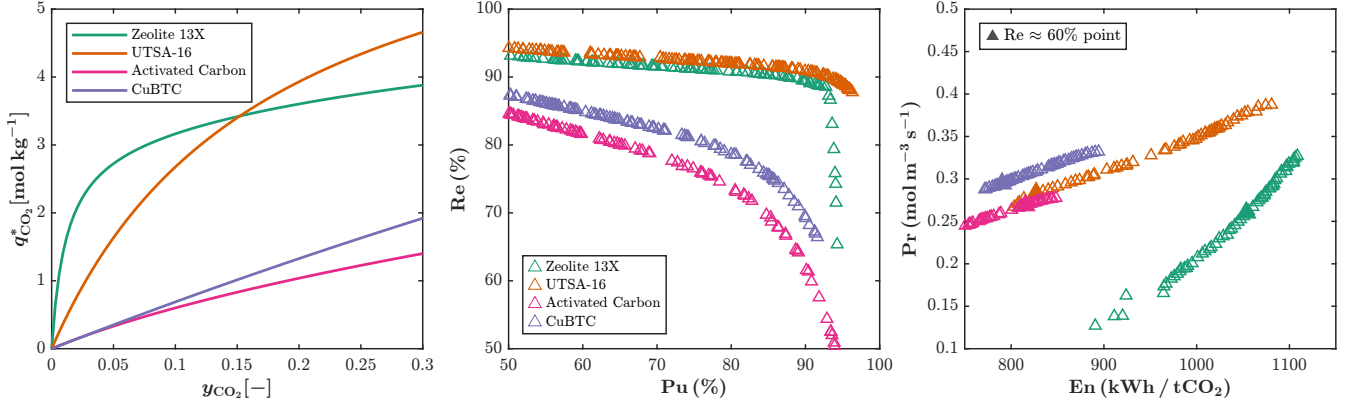


Fig. 5. KPI comparison for selected adsorbents: (a) Isotherms, (b) Pu-Re, (c) En-Pr, with solid triangles denoting operating points where recovery is around 60%.

and low energy usage. Despite CuBTC and Activated Carbon both exhibiting high k_1 , insufficient Δq_{CO_2} and $S_{ads, CO_2/N_2}$ limit both optimal purities to $\sim 80\%$. Conversely, Mg-MOF-74, despite exceptional Δq_{CO_2} , cannot reach $Pu \geq 80\%$ because slow uptake and poor CO_2/N_2 discrimination prevent clean product formation.

3.4.2. Cross comparison of KPIs

Four adsorbents, Zeolite 13X, UTSA-16, CuBTC, and Activated Carbon, are selected for their varied k_1 values and isotherm properties to explore the impact of k_1 and q^* on the KPIs. Figure 5(a) shows the isotherms for the 4 materials, while (b) and (c) depict PFs on the Pu-Re and En-Pr domains, respectively.

Zeolite 13X and UTSA-16 show strongly curved CO_2 isotherms, leading to abrupt bends on the Pu-Re PFs. While both materials have strong isotherms providing excellent CO_2 separation, the regeneration shown by t_{ads} and t_{evac} is slow due to poor kinetics. Consequently, they achieve high Pu and Re at the cost of high En_{min} and low Pr_{max} . In contrast, Activated Carbon and CuBTC have weaker isotherms but much better mass-transfer kinetics. As a result, they deliver poorer separation but are significantly more energy efficient.

It is evident that En_{min} is driven by $S_{ads, CO_2/N_2}$ and Δq_{CO_2} alongside k_1 ; and fast kinetics do not guarantee high Pu or Re. Top performers balance Δq_{CO_2} , $S_{ads, CO_2/N_2}$, and k_1 whereas materials exhibiting excellent property in only one dimension fall inside the Pareto frontier when considering full-cycle behaviours.

3.5. Mapping of ideal adsorbent

An ideal adsorbent would exhibit high k_1 , and $S_{ads, CO_2/N_2}$. To construct such an adsorbent, the k_1

value of Zeolite 13X was arbitrarily perturbed by additions of $0.06 [s^{-1}]$ and $0.12 [s^{-1}]$. The manipulated k_1 , Pu-Re PFs, and En-Pr PFs are shown in Figure 6(a)(b)(c), respectively. The Pu-Re domain is markedly insensitive to the changes, where all three cases (baseline k_1 , $k_m + 0.06$, and $k_m + 0.12$) collapse onto nearly identical Pu-Re PFs.

Pr, by contrast, shows the strong kinetic response. The increase of k_1 has shifted the En-Pr PF upward, delivering a substantial rise in Pr at broadly unchanged En range. Diminishing returns are apparent. The increase from baseline k_1 to $k_m + 0.06$ produces significant productivity gains. The second increase to $k_m + 0.12$, however, yields only a marginal shift. It can be concluded that Pr responds non-linearly, improving rapidly until a threshold is reached, after which the cycle becomes equilibrium-controlled.

This validates the significance of the k-BAAM model for its ability to reflect the effect of kinetics on KPIs. The same conclusion could not be drawn from the BAAM model as it does not consider kinetics.

Additionally, this feature of the k-BAAM model can serve as a guide for material design.

4. Conclusion

This work used a kinetics-augmented BAAM (k-BAAM) that simulates a PVSA cycle and applied it as a shortcut tool for screening sorbents for point-source CO_2 capture. By relaxing the instant equilibrium assumption and embedding an LDF-driven mass-transfer coefficient k_1 into the BAAM framework, k-BAAM accounts for

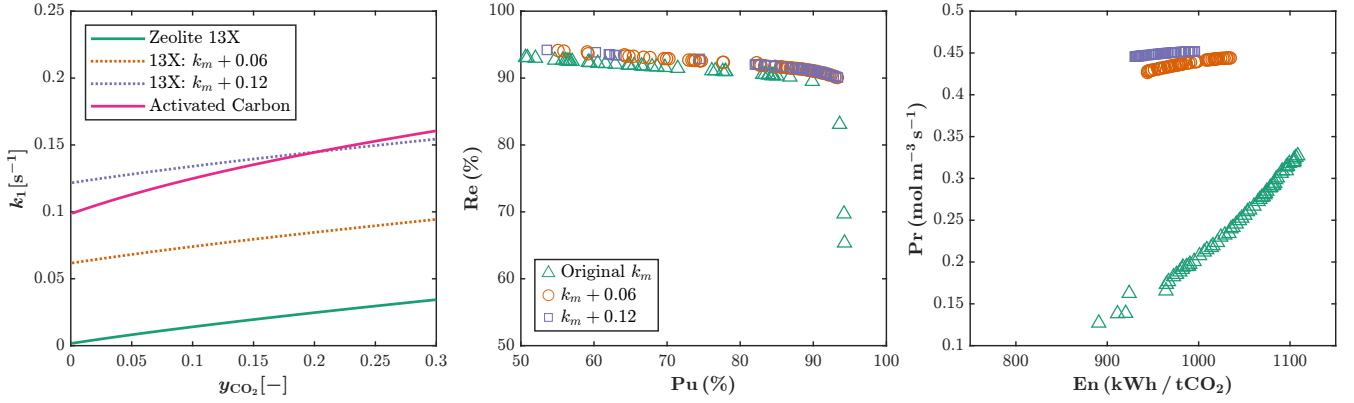


Fig. 6. Effect of increasing k_1 on Zeolite 13X performance: (a) arbitrary change of k_1 , (b) Pu-Re, (c) En-Pr

time derivatives and adsorption kinetics while preserving the computational efficiency of the original equilibrium BAAM model. At a run time of 5sec per data, the k-BAAM is 60x faster than the HiFi Model, successfully reducing 98% of computational time. It provides physically reliable estimates of key KPIs—purity, recovery, energy and productivity—across a broad design space, thereby bridging the gap between the simplistic BAAM model and the computationally intensive High-Fidelity model.

Of the 7 adsorbents examined—Zeolite 13X, UTSA-16, CALF-20, MIL-160(Al), CuBTC, Activated Carbon, and Mg-MOF-74—UTSA-16 yields the best overall performance with highest purity, recovery, and productivity. Equilibrium descriptors of the materials—isorthem value, working capacity, and selectivity—impact the separation of CO₂ while adsorption kinetics—mass transfer efficiency—governs the performance of productivity and energy consumption. Top-performing adsorbents would balance strong isotherms, high selectivity, and good kinetics, validating the significance and necessity of the kinetic approach.

5. Outlook

The current k-BAAM model is immediately applicable to screen existing and emerging materials for PVSA-based point-source CO₂ capture. Its low computational cost and credibility allow routine evaluation of large sorbent libraries. Explorations of operating conditions for any binary CO₂/N₂ PVSA configuration compatible with all assumptions can also be conducted immediately. The model could also provide guidance for materials development, as k-BAAM incorporates adsorption kinetics, a factor with direct correlations to the physical properties of materials. The feasibility of the changes in material production, however, is to be decided. Several extensions would further increase the scope and usage of the k-BAAM. Firstly, an increase in desorption pressure p_L could be investigated to understand the effect of p_L on specific energy. Secondly, the present study focused on low to moderate point-source feed compositions. Extension of analysis to more dilute regimes, i.e. 3% CO₂ feed concentration, would help assess the ap-

plicability of the model in more versatile carbon capture technologies. Thirdly, the optimisation boundaries can be relaxed or tightened to further understand the KPI performances. Key characteristics discussed— k_1 , q^* , and $S_{ads,CO_2/N_2}$ —can also be explored to further understand the individual effect on KPIs. Furthermore, coupling k-BAAM outputs with simple economic metrics such as cost of energy alongside CapEx and OpEx KPIs would enable more holistic decision-making. Lastly, incorporating non-isothermal behaviour into k-BAAM would enable quantification of the impact of adsorption heat and thermal management on the PFs.

In the longer term, k-BAAM could be embedded within a multi-stage screening pipeline that combines equilibrium BAAM, k-BAAM, and the High-Fidelity model. In such a workflow, BAAM would act as an initial equilibrium filter; k-BAAM would provide a refinement of results under the consideration of kinetics; the High Fidelity model would be reserved for top candidates to detail precise behaviours. The integration would provide a flexible sorbent-screening environment capable of supporting both academic studies and industrial development of CO₂ capture processes.

Acknowledgements

The authors thank Professor Ronny Pini and Elena Chastali for the support provided throughout the research.

References

- 1) IPCC: *Climate Change 2021: the Physical Science Basis*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2023 doi:10.1017/9781009157896.
- 2) International Energy Agency: *Direct Air Capture*, IEA Publications, Paris, 2022.
- 3) Ozkan, M. et al.: Current Status and Pillars of Direct Air Capture Technologies, *iScience*, **25**, 4 (2022), 103990 doi:10.1016/j.isci.2022.103990.
- 4) Erans, M. et al.: Direct Air Capture: Process Technology, Techno-economic and Socio-political Challenges, *Energy Environmental Science*, **15** (2022), pp. 1360–1405 doi:10.1039/D1EE03523A.
- 5) McQueen, N. et al.: A Review of Direct Air Capture (DAC): Scaling up Commercial Technologies and Innovating for the Future, *Progress in Energy*, **3**, 3 (2021), 032001 doi:10.1088/2516-1083/abf1ce.

- 6) Krevor, S. et al.: Subsurface Carbon Dioxide and Hydrogen Storage for a Sustainable Energy Future, *Nature Reviews Earth Environment*, **4**, 2 (2023), pp. 102–118 doi:10.1038/s43017-022-00376-8.
- 7) Artz, J. et al.: Sustainable Conversion of Carbon Dioxide: an Integrated Review of Catalysis and Life Cycle Assessment, *Chemical Reviews*, **118**, 2 (2017), pp. 434–504 doi:10.1021/acs.chemrev.7b00435.
- 8) Liu, C. et al.: A Life Cycle Assessment of Greenhouse Gas Emissions from Direct Air Capture and Fischer–Tropsch Fuel Production, *Sustainable Energy Fuels*, **4**, 6 (2020), pp. 3129–3142 doi:10.1039/c9se00479c.
- 9) Deutz, S. and Bardow, A.: Life-cycle Assessment of an Industrial Direct Air Capture Process Based on Temperature–vacuum Swing Adsorption, *Nature Energy*, **6**, 2 (2021), pp. 203–213 doi:10.1038/s41560-020-00771-9.
- 10) Zhu, X. et al.: Recent Advances in Direct Air Capture by Adsorption, *Chemical Society Reviews*, **51**, 15 (2022), pp. 6574–6651 doi:10.1039/d1cs00970b.
- 11) Balashankar, V., Rajagopalan, A., Pauw, R., Avila, A., and Rajendran, A.: Analysis of a Batch Adsorber Analogue for Rapid Screening of Adsorbents for Postcombustion CO₂ Capture, *Industrial Engineering Chemistry Research*, **58**, 8 (2019), pp. 3314–3328 doi:10.1021/acs.iecr.8b05420.
- 12) Ward, A. and Pini, R.: Efficient Bayesian Optimization of Industrial-Scale Pressure–Vacuum Swing Adsorption Processes for CO₂ Capture, *Industrial Engineering Chemistry Research*, **61**, 36 (2022), pp. 13650–13668 doi:10.1021/acs.iecr.2c02313.
- 13) Ward, A. and Pini, R.: Design and Performance Evaluation of Multisorbent Vacuum-Swing Adsorption Processes for Postcombustion Carbon Capture, *Industrial Engineering Chemistry Research*, **63**, 31 (2024), pp. 13787–13800 doi:10.1021/acs.iecr.4c01345.
- 14) Peh, S., Farooq, S., and Zhao, D.: A Metal-organic Framework (MOF)-based Temperature Swing Adsorption Cycle for Postcombustion CO₂ Capture from Wet Flue Gas, *Chemical Engineering Science*, **250** (2021), 117399 doi:10.1016/j.ces.2021.117399.
- 15) Lin, J.-B. et al.: A Scalable Metal-organic Framework as a Durable Physisorbent for Carbon Dioxide Capture, *Science*, **374**, 6574 (2021), pp. 1464–1469 doi:10.1126/science.abi7281.
- 16) Ward, A., Papathanasiou, M., and Pini, R.: The Impact of Design and Operational Parameters on the Optimal Performance of Direct Air Capture Units Using Solid Sorbents, *Adsorption*, **30**, 7 (2024), pp. 1829–1848 doi:10.1007/s10450-024-00526-y.
- 17) Pai, K., Prasad, V., and Rajendran, A.: Practically Achievable Process Performance Limits for Pressure–Vacuum Swing Adsorption-Based Postcombustion CO₂ Capture, *ACS Sustainable Chemistry Engineering*, **9**, 10 (2021), pp. 3838–3849 doi:10.1021/acssuschemeng.0c08933.
- 18) Farmahini, A. et al.: Performance-Based Screening of Porous Materials for Carbon Capture, *Chemical Reviews*, **121**, 17 (2021), pp. 10666–10741 doi:10.1021/acs.chemrev.0c01266.

Reaction Modelling and Geometry Optimisation for Multi-tubular Membrane Catalytic Ammonia Cracker

Ziheng Zhao

Department of Chemical Engineering,
Imperial College London,
London, SW7 2AZ, UK
email: ziheng.zhao22@imperial.ac.uk

This study develops and optimises a multi-tubular catalytic membrane reactor (CMR) for a small-scale hydrogen synthesis process via ammonia cracking, which will potentially be employed in the transportation sector for hydrogen fuel utilisation. A coupled reaction–separation equilibrium model is formulated to capture ammonia thermal decomposition enhanced by continuous hydrogen removal through palladium membrane, enabling high conversion with milder operating conditions than the conventional reverse Haber–Boch process. The model included detailed kinetics pathways and mass transfer behaviour. The model simulation result was validated against literature trends. To improve reactor compactness and catalyst utilisation, a multi-objective optimisation framework based on the NSGA-II evolutionary algorithm is implemented using five design variables: inner diameter, outer diameter, reactor length, inlet temperature, and inlet pressure. The resulting Pareto front reveals clear trade-offs between volumetric hydrogen productivity and catalyst-specific productivity, highlighting two knee regions that correspond to optimal geometries for volume-constrained versus cost-sensitive applications. Geometric trends show that long, slender reactors maximise H_2 yield per volume, whereas shorter, wider designs favour catalyst efficiency. Sensitivity analyses on temperature, pressure, and ammonia inlet flow confirm that high conversion is strongly temperature-dependent, and that high flow rates introduce dynamic control challenges despite achieving similar final conversions. Overall, the work demonstrates that combining mechanistic modelling with multi-objective optimisation provides a solid framework for designing a high-performance multitubular catalytic membrane reactor for ammonia decomposition, with potential applications for next-generation hydrogen-based transport systems.

Keywords: Ammonia thermal decomposition, Hydrogen production, Membrane reactor design, NSGA-II Optimisation

1 Introduction

1.1 Background. The global transition toward a low-carbon energy system has placed significant pressure on all major sectors to reduce greenhouse gas emissions. Among these, the transportation sector remains particularly challenging to decarbonise due to its dependence on high-energy-density liquid fuels and complex logistical requirements [1]. Aviation and maritime transport in particular represent two hard-to-abate modalities, where direct electrification remains technologically and economically constrained [2].

To meet the energy demands of long-distance transport while reducing emissions, the industry has turned to hydrogen as a clean alternative fuel. Hydrogen offers near-zero carbon emissions at the point of use and can be produced sustainably through renewable-powered electrolysis, positioning it as a key enabler of future net-zero transport systems. Its high specific energy and compatibility with emerging technologies such as fuel cells make hydrogen particularly attractive for aviation and maritime applications where onboard energy requirements are substantial. Nevertheless, practical deployment remains constrained by the difficulties of storing and transporting hydrogen due to its low volumetric energy density, cryogenic or high-pressure storage requirements, and safety concerns related to leakage and flammability [3,4]. These challenges have prompted growing interest in chemical carriers that can store hydrogen more efficiently and safely.

Ammonia has recently emerged as a promising energy vector capable of supporting large-scale hydrogen utilisation. As a carbon-free molecule with high volumetric hydrogen density, ammonia offers significant advantages in storage and distribution compared to molecular hydrogen. In particular, the use of ammonia mitigates key risks associated with hydrogen storage, including hydrogen embrittlement, leakage, and explosion hazards due to hydrogen's high diffusivity and volatility [5,6]. These characteristics have motivated substantial research into ammonia cracking technologies, enabling on-demand hydrogen production for fuel cells and combustion systems through the reverse Haber–Bosch (ammonia decomposition) reaction pathway. A persistent challenge lies in removing residual ammonia and nitrogen from product mixtures to meet purity requirements for downstream applications. Palladium-based membranes are of particular interest due to their exceptional hydrogen selectivity and permeability [7,8]. Integrating membranes into catalytic membrane reactors (CMRs) allows simultaneous reaction and separation, improving equilibrium conversion and hydrogen yield [9]. Multiple system-level studies have demonstrated the potential of membrane-assisted ammonia cracking to achieve compact, energy-efficient designs suitable for decentralised and mobility-oriented application of ammonia as hydrogen fuel stock in transportation sectors [10,11].

Building upon the CMR (catalytic membrane reactor) design framework reported by Guo et al. [12], there are further opportunities to enhance reactor performance through rigorous

optimisation in its geometric structure. In aviation and maritime applications, reactor mass and volume are critical engineering constraints; thus, reducing reactor volume and catalyst inventory without compromising hydrogen recovery is essential. This motivates the formulation of a multi-objective optimisation problem focused on minimising both reactor volume and catalyst requirement while ensuring adequate membrane hydrogen permeation. The optimisation considers four decision variables — namely, inlet temperature, inlet pressure, reactor diameter, and reactor length—each bounded by physically meaningful limits. Additionally, a minimum hydrogen permeation flux is imposed to maintain system feasibility.

To address this optimisation challenge, the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) is employed. NSGA-II is a widely adopted evolutionary optimisation method known for efficiently exploring complex, nonlinear, multi-modal design spaces [12–14]. Its strengths include fast non-dominated sorting, diversity preservation via crowding distance, and effective convergence towards well-distributed Pareto-optimal fronts without requiring gradient information. These properties make NSGA-II particularly suitable for engineering design problems involving conflicting objectives, such as simultaneously minimising reactor size and catalyst usage while maximising hydrogen recovery.

The optimised reactor configurations are subsequently validated through sensitivity analyses examining ammonia conversion across a range of temperatures and pressures. This ensures that the solutions not only satisfy optimisation criteria but also remain consistent with the underlying thermodynamic and kinetic behaviour of ammonia decomposition.

1.2 Research Scope and Contribution. This study builds upon the multitubular catalytic membrane reactor developed by Guo et al [15], by shifting the focus from preliminary feasibility to the optimisation of its mechanical design for improved industrial and mobility applications. A multi-objective optimisation framework is developed to minimise reactor volume, weight, and catalyst requirement, using inlet temperature, inlet pressure, reactor diameter, and reactor length as key design variables subject to engineering constraints.

A major contribution of this work is the application of the NSGA-II algorithm to navigate the complex, nonlinear design space of a reaction–separation membrane reactor. This study demonstrates the suitability of NSGA-II for mechanical design and optimisation, it also highlights how machine-learning-based optimisation techniques can be integrated with traditional engineering workflows. The resulting Pareto-optimal solutions provide valuable insights into design trade-offs and offer a more compact and efficient reactor configuration compared to the preliminary design.

Overall, this research showcased the broader use of evolutionary optimisation methods in reactor and mechanical system design.

2 Model Development

2.1 Physical description. As shown in Fig. 1, a compact multitubular membrane reactor is investigated in this study. The reactor is divided by the membrane into the reaction zone and the separation zone [16]. Ammonia decomposition, as in Eq. 1, is the main reaction of the reactor, and the side reactions can be considered negligible. In the configuration shown, ammonia (NH_3) along with a small amount of water vapour enters the reaction zone, where it is continuously converted into hydrogen (H_2) and nitrogen (N_2) with the aid of heat and a $\text{Ni}/\text{Al}_2\text{O}_3$ catalyst. At the same time, the

generated hydrogen permeates through the palladium membrane into the separation zone due to the pressure gradient across the membrane, allowing for both product purification and a favourable shift in reaction equilibrium.

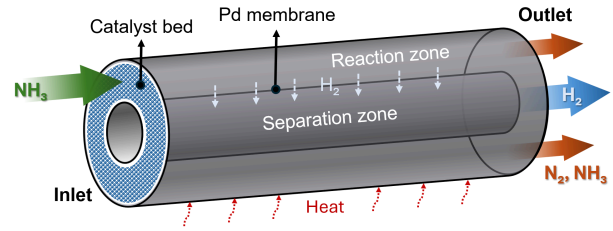


Fig. 1 Schematic diagram of multitubular membrane reactor.

2.2 Model Description. To evaluate the integrated reaction and separation processes occurring within the reactor, a rapid and efficient computational model is developed. Because this study aims to optimise the reactor geometry once equilibrium is achieved, a model with reduced computational demands is preferable. Accordingly, a thermodynamic equilibrium framework is employed. The model formulation is based on the following assumptions:

- (1) Temperature, pressure, and flow within the reactor are uniform and constant.
- (2) The hydrogen separation membrane has 100% selectivity.
- (3) The system operates under isothermal conditions.
- (4) The gas phase is perfectly mixed, such that radial diffusion and concentration polarisation are neglected.

The kinetic behaviour of ammonia thermal decomposition is modelled using the Temkin-Pyzhev mechanistic rate expression [17], as shown in Eq. 2, 3 and 4:

$$r_{\text{NH}_3} = \eta \cdot k \left[\left(\frac{p_{\text{NH}_3}^2}{p_{\text{H}_2}^3} \right)^\beta - \frac{p_{\text{N}_2}}{k_{eq}^2} \left(\frac{p_{\text{H}_2}^3}{p_{\text{NH}_3}^2} \right)^{1-\beta} \right] \quad (2)$$

$$k = 5.744 \times 10^{19} \cdot \exp \left[\frac{-2.304 \cdot 10^5}{T} \right] \quad (3)$$

$$\frac{1}{k_{eq}} = \exp \left[\frac{2250.322}{T} - 0.8534 - 1.51049 \cdot \ln T - 2.58987 \times 10^{-4} \cdot T + 1.48961 \times 10^{-7} \cdot T^2 \right] \quad (4)$$

Where r_{NH_3} denotes the ammonia decomposition rate ($\text{mol} \cdot \text{s}^{-1} \cdot \text{m}^{-3}$), p represents the partial pressure of each species (Pa), k is the reaction rate constant, and β is a dimensionless parameter with a value of 0.674 [17]. To obtain a more accurate description of the reaction behaviour, a combined Knudsen and binary diffusion framework is used to determine the effective diffusivity. This approach incorporates molecular interactions within narrow pores, leading to improved prediction of the reaction rate constant. The molecular interaction parameter, expressed as Φ_{ij} , is evaluated by Eq.5 [4,15]:

$$\Phi_{ij} = \frac{1}{\sqrt{8}} \left[1 + \frac{M_i}{M_j} \right]^{-1/2} \left[1 + \sqrt{\frac{\mu_i}{\mu_j}} \left(\frac{M_j}{M_i} \right)^{1/4} \right]^2 \quad (5)$$

Eq. 6 determines the mixture viscosity (μ_{mix}) by combining the viscosities of the individual gas species (μ_i) according to their respective mole fractions. The calculation also incorporates the interaction parameters obtained in the preceding step, allowing for a more accurate estimation of the overall gas mixture viscosity.

$$\mu_{mix} = \sum_i \frac{x_i \mu_i}{\sum_{j \neq i} \Phi_{ij} x_j} \quad (6)$$

Eq. 7 evaluates the overall molecular interaction parameter ($\sigma_{mix}^2 \omega_{mix}$) for the gas mixture. It incorporates the mixture molecular weight (M_{mix}), temperature (T), and mixture viscosity (μ_{mix}) to represent the strength of intermolecular forces and gas-phase interactions.

$$\sigma_{mix}^2 \omega_{mix} = \frac{2.6693 \times 10^{-5} - \sqrt{M_{mix} T}}{\mu_{mix}} \quad (7)$$

Eq. 8 determines the binary diffusion coefficient (D_{mn}) between two gaseous species. The correlation accounts for the system temperature (T), pressure (P), molecular weights (M_m), and the previously calculated molecular interaction term ($\sigma_{mix}^2 \omega_{mix}$), providing a more accurate representation of gas-phase diffusivity.

$$D_{mn} = 0.0018583 \sqrt{T^3 \left(\frac{1}{M_m} + \frac{1}{M_n} \right) \frac{1}{P \sigma_{mix}^2 \omega_{mix}}} \quad (8)$$

From this point, the Knudsen diffusion coefficient is evaluated with the following procedure. Eq. 9 provides the formulation for determining the Knudsen diffusion coefficient, as reported in [13]:

$$D_{k,m} = \frac{2}{3} r_{av} \sqrt{\frac{8RT}{\pi M_{mix}}} \quad (9)$$

The effectiveness factor, η , as used in Eq. 2, is determined by incorporating both the Knudsen and binary diffusion coefficients. The evaluation proceeds through several steps, beginning with the introduction of an intermediate parameter, $D_{p0,m}$, defined in Eq. 10:

$$D_{p0,m} = \left[\frac{1}{D_{mn}} + \frac{1}{D_{k,m}} \right]^{-1} \quad (10)$$

Eq. 11 determines the effective diffusion coefficient (D_{pm}) within the porous catalytic bed. It modifies the intermediate diffusion coefficient ($D_{p0,m}$) by accounting for the catalyst bed void fraction (ϵ), solid catalyst density (ρ_{solid}), and tortuosity (τ_{solid}), thereby capturing the realistic structural and transport behaviour of the catalyst medium.

$$D_{p,m} = D_{p0,m} \frac{\epsilon \rho_{solid}}{\tau} \quad (11)$$

The Thiele modulus (h_T), a dimensionless number describing the balance between reaction rate and pore diffusion rate, is defined in Eq. 12. It incorporates the catalyst particle radius (r_p), the reaction kinetic parameter (K_v), and the effective diffusion coefficient (D_{pm}).

$$h_T = \frac{d_p}{6} \sqrt{\frac{k (RT/P) \rho_p}{D_{pm}}} \quad (12)$$

The internal effectiveness factor is thus calculated from Eq. 13.

$$\eta = \frac{3}{h_T} \left[\coth(h_T) - \frac{1}{h_T} \right] \quad (13)$$

The axial mass balance in the reaction zone is formulated as shown in Eq. 14, where γ_i represents the stoichiometric coefficient for each species (NH_3 , H_2 , and N_2). The permeation rate of each species through the membrane is expressed as Q_i .

$$\frac{d\dot{n}_i}{dL} = \gamma_i \cdot r \cdot A_{cs} - Q_i \quad (14)$$

According to the model assumption, only hydrogen permeates through the palladium membrane. The permeation rate of H_2 through the palladium membrane is mathematically expressed by Eq. 15 [16].

$$Q_{H_2} = 0.223 \exp \left[\frac{-1903}{T} \right] \left(p_{H_2}^{0.5} - p_{H_2,in}^{0.5} \right) \pi D_2 \quad (15)$$

The pressure drop along the axial direction of the reaction zone under fully developed turbulent flow is described using the Ergun equation, as shown in Eq. 16. Furthermore, with the isothermal assumption applied, the temperature remains constant throughout the reactor length, leading to the simplified expression in Eq. 17.

$$\frac{dP}{dL} = - \frac{1.8 (1 - \epsilon) \rho_g}{d_p \epsilon^3} \left(\frac{u_0^2}{A_{cs} \rho_p} \right) \quad (16)$$

$$\frac{dT}{dL} = 0 \quad (17)$$

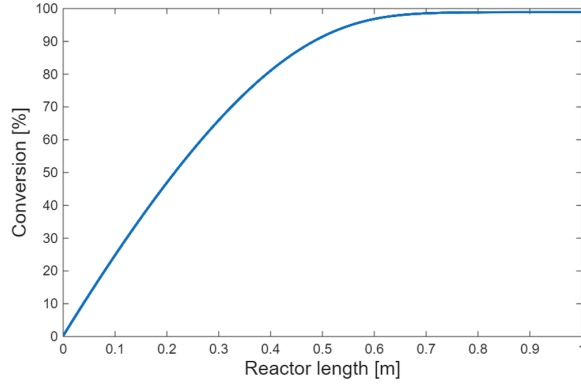
A system of ordinary differential equations (ODEs) with respect to reactor length is obtained from combining Eq. 14, 16 and 17. These ODEs are solved using the solve_ivp numerical method in SciPy (Python 3). Python offers efficient numerical libraries and flexible code integration, making it well-suited for advanced engineering design and analysis applications.

2.3 Model Validation. To evaluate the model performance and justify the need for optimisation, a representative set of reactor parameters is selected, as listed in Table 1. The predicted ammonia conversion along the reactor length is shown in Fig. 2(a). The results indicate that a very high conversion level (>98.7%) can be achieved when the reactor is sufficiently long, even under relatively mild operating conditions of 450 °C and 3 bar. This is notable, as conventional thermal cracking typically requires temperatures of at least 500 °C. The improved conversion arises from equilibrium enhancement, where continuous hydrogen permeation through the membrane to the separation zone shifts the reaction equilibrium toward further ammonia decomposition, thereby increasing the overall hydrogen production.

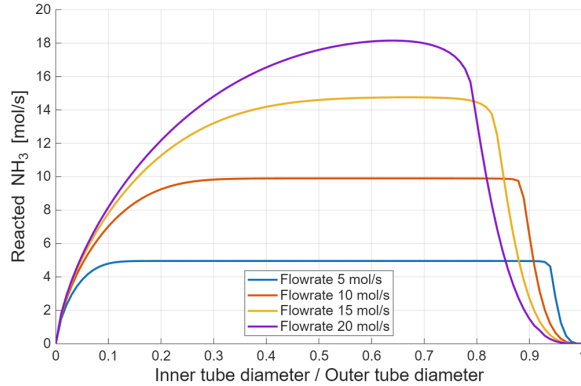
The graph Fig. 2(b) shows how, for a given ratio of inner to outer tube diameter, the rate of ammonia conversion depends on the value of the inlet flowrate. Thus, for low flowrates, ammonia conversion increases quickly from both ends of extreme diameter ratios, and then levels out as geometry differences are no longer important when a sufficient volume of catalyst and permeation area are guaranteed. However, for higher flowrates, the curves exhibit a clear maximum, where improvement in ammonia conversion with increasing inner diameter reflects enhanced permeation, but thereafter decreases as the reaction zone is no longer thick enough to provide sufficient catalyst volume and optimal reaction conditions. This shows how, at high flowrate operation, geometrical differences in the reactor cause a large, non-linear affect, with some geometries being rather superior to others. This proves that an optimisation on reactor geometry and condition is crucial for a high conversion.

Configuration	Value
Outer diameter [m]	0.1
Inner diameter [m]	0.06
Reactor tube length [m]	1
Ammonia inlet flowrate [mol/s]	10
Operating temperature [°C]	457
Operating pressure [MPa]	3.0
Inner tube pressure [Pa]	2000
Catalyst pack bed voidage	0.6
Tortuosity factor (τ)	4
Catalyst particle radius (r_p) [m]	3×10^{-4}
Catalyst bulk density (ρ_{solid}) [kg/m ³]	1200

Table 1 Representative configuration used for validation.



(a) NH_3 conversion profile along the reactor length.



(b) The impact of NH_3 flowrate on reacted NH_3 at different inner to outer diameter ratio.

Fig. 2 Reactor model simulation results.

3 Geometry Optimisation

3.1 Optimisation Problem Definition. An optimisation study is performed to maximise (i) the hydrogen permeation rate per reactor volume and (ii) the hydrogen permeation rate per catalyst mass, thereby assessing the enhancement potential of the membrane reactor. The first objective reflects the volumetric productivity of the system, whereas the second indicates how efficiently the catalyst is utilised under identical feed conditions. Both metrics are essential for transportation applications, although their relative importance varies by sector. For instance, aviation places strict constraints on equipment volume and may therefore prioritise volumetric productivity, while maritime applications typically involve significantly higher fuel consumption. In such cases, minimising catalyst usage – due to its high cost – could be more critical than reducing reactor volume.

The optimised design and operating variables in this study include the inner diameter of the inner tube d_{in} , the diameter of the outer tube d_{out} , and the reactor length L , which together determine the geometric configuration and membrane surface area available for hydrogen extraction. In addition, the inlet temperature T_{in} and inlet pressure P_{in} were selected as decision variables to capture the influence of thermodynamic driving forces and reaction kinetics on system performance. These five optimisation variables collectively govern the mass transfer, reaction rate, and separation efficiency within the membrane reactor.

The optimisation of these design and operating conditions was conducted using the second version of the non-dominated sorting genetic algorithm (NSGA-II), selected due to its robust handling of complex, non-linear, and multi-objective problems as well as the ease with which it integrates with the developed reactor model. NSGA-II efficiently explores trade-offs between conflicting performance objectives while respecting operational constraints. The generic formulation of the optimisation problem is expressed as:

$$\text{Minimise: } \mathbf{F}(\mathbf{x}) = \begin{bmatrix} \dot{n}_{H_2}/V_{\text{reactor}}(\mathbf{x}) \\ \dot{n}_{H_2}/m_{\text{cat}}(\mathbf{x}) \end{bmatrix} \quad (18)$$

$$\text{Subject to: } \begin{cases} 0.01 \leq d_{in} \leq 0.15 & [\text{m}] \\ 0.01 \leq d_{out} \leq 0.15 & [\text{m}] \\ 0.05 \leq L \leq 1.00 & [\text{m}] \\ 530.0 \leq T_{in} \leq 730.0 & [\text{K}] \\ 0.1 \leq P_{in} \leq 3.0 & [\text{MPa}] \end{cases} \quad (19)$$

$$\text{with } \mathbf{x} = [d_{in}, d_{out}, L, T_{in}, P_{in}] \quad (20)$$

Other types of constraints are incorporated into the NSGA-II optimisation through penalty terms applied to the objective evaluations. Two conditions were treated as fatal constraints: (i) when the hydrogen permeation exceeded the specified upper limit, and (ii) when the ODE solver failed to converge for a given set of operating and geometric variables. In either case, a very large penalty is added to the objective values, effectively removing these solutions from consideration. In addition, a soft constraint was applied for the reactor conversion requirement. When the outlet conversion fell below the target value of 99%, a penalty proportional to the difference from the target is introduced. The proportionality factor is tuned so that the order of magnitude of this conversion penalty is comparable to that of the normalised objective functions, allowing the optimisation process remains sensitive to the conversion without overwhelming the objectives. This hybrid penalty strategy lets the algorithm to explore the design space widely while discouraging infeasible operating conditions.

3.2 Multi-objective Optimisation. The NSGA-II algorithm begins by initialising a set of random decision variables. These variables are taken to the equilibrium model for calculation. Next, the algorithm evaluates each calculated result and performs ranking using fast non-dominated sorting to identify Pareto fronts. After ranking, NSGA-II selects the best individuals for reproduction by prioritising those with lower rank while considering the crowding distance to explore more possibilities. These selected individuals then reproduce with crossover and mutation, simulating evolution

of life and creating new offspring decision variables. The new offspring are once again combined with the parent population, and the process repeats to gradually converge towards an optimised set of decision variables. Each full cycle of evaluation, ranking, selection, and reproduction in the evolutionary process is called a generation [?]. A route map is shown in Fig. 3 to illustrate the process. In this research, the initial population of size is set at 50 and the termination condition is defined as either the maximum number of generations reaches 50, or the number of stalled generations reaches 20.

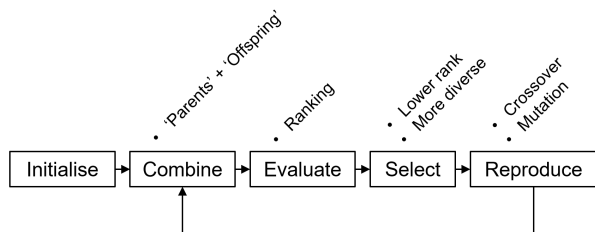


Fig. 3 Flow chart of NSGA-II optimisation algorithm.

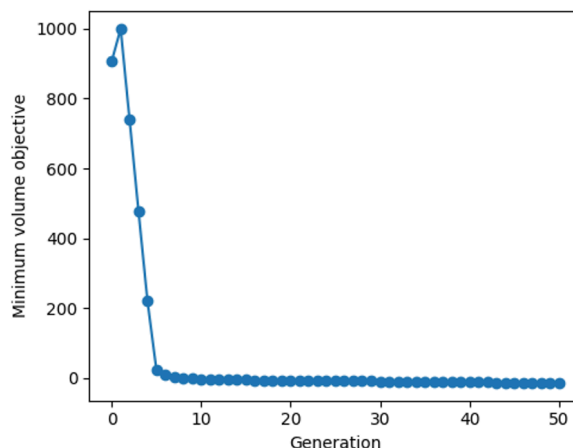
To ensure a fair and meaningful optimisation, both objective functions are normalised so that they have comparable influence in the numerical solver. Since these metrics have inherently different units and magnitudes, direct multi-objective evaluation would bias the optimisation toward whichever metric dominates numerically. Therefore, before running optimisation each objective was independently normalised using its respective maximum value obtained from preliminary simulations. Such process brings the two performance indicators onto a common scale between 0 and 1, ensuring neither objective overshadows the other during trade-off evaluation. As a result, equal weighting in the multi-objective framework genuinely reflects equal design priority between maximising hydrogen extraction efficiency per reactor volume and per catalyst utilisation.

Following normalisation, a weighting coefficient was applied to each objective function to further balance its contribution relative to the penalty terms introduced for constraint handling. This coefficient ensures that maximising both objectives remains the dominant driver in the optimisation, while still allowing operational and design constraints to be adequately enforced. By tuning these coefficients, the optimisation algorithm maintains numerical stability and avoids scenarios where penalties either overwhelm the performance objectives or become insignificant. This scaling approach ensures that both objectives and penalties coexist within a common and physically meaningful range, enabling robust and unbiased identification of optimal reactor configurations. It also allows changes of objectives weighting in further research.

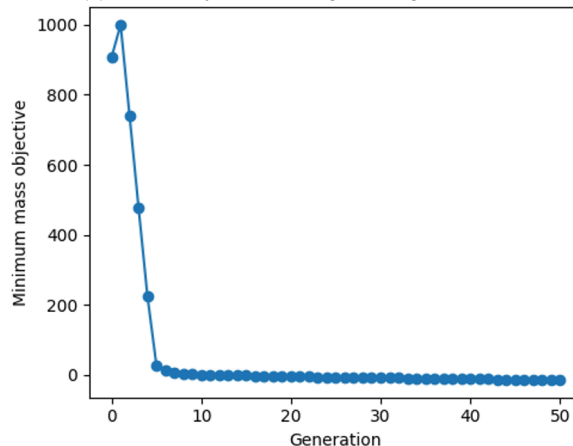
The calculations are carried out on a personal laptop with a 20-core parallel 13th Gen Intel(R) Core(TM) i7-13650HX CPU and 16 Gb RAM.

3.3 Optimisation Results and Discussion. Fig. 4 represents the two objective functions – hydrogen permeation per unit catalyst mass and per unit reactor volume over 50 generations of NSGA-II. A rapid drop of the objective values in the first 5-10 generations shows that the algorithm is extremely efficient at identifying feasible designs with significantly improved hydrogen productivity. For generations beyond about the 10th, both curves start to level off, reflecting solutions approaching a near-optimum, stable Pareto region. No further improvement can be seen beyond the 30th and 34th generation respectively for each objective, hence optimum to the limit of the algorithm has been achieved. It has

to be noticed that the objectives are plotted as negative values, because NSGA-II is a minimisation-based algorithm, hence maximising hydrogen productivity was implemented in the code by minimising the negative of these performance metrics. After taking the opposite number, the penalties are added. Therefore, more negative values correspond to more favourable reactor designs. The average decision variables of the final population is reported in Table 2. The temperature and the pressure always reach the upper boundary, which makes sense as the reaction itself favours high temperature, and higher pressure allows more hydrogen permeation across the membrane and further drive the reaction equilibrium forward.



(a) Volume objective converges over generations



(b) Mass objective converges over generations

Fig. 4 The convergence of the two objective functions with penalties included.

Decision variable	Value
Outer diameter [m]	0.1085
Inner diameter [m]	0.0810
Reactor tube length [m]	0.5602
Operating temperature [K]	730.0 (Upper boundary)
Operating pressure [MPa]	3.0 (Upper boundary)

Table 2 Average optimised value of decision variables.

Fig. 5 presents the final Pareto front obtained from the NSGA-II optimisation, illustrating the trade-off between hydrogen permeation per unit catalyst mass and per unit reactor volume. The generally smooth, continuous distribution of non-dominated points proves that NSGA-II is able to properly explore the feasible region and converge to a well-defined set of optimal solutions. The

overall curvature of the front reflects the progressive compromise between volumetric and catalyst-specific hydrogen productivity – gains in one objective necessarily come at the expense of the other, and the slope of the curve is an indicator of the sensitivity of each metric to such a trade-off. Regions characterised by a slope correspond to regimes where a gain in one objective can be achieved with relatively low penalties in the other, while at both extremes of slope, either very steep or very flat, reflects poor exchange rates between the objectives, meaning one must give up a lot in one metric to gain very little in the other.

Within this frontier, two distinct knee regions are apparent, marked by great changes in slope. Both knees represent extreme points – one nearly vertical and the other nearly horizontal. The upper knee corresponds to designs that achieve exceptionally high hydrogen productivity per unit catalyst mass but only moderate volumetric efficiency; beyond this point, further improvements in catalyst-specific performance become nearly unattainable. Conversely, the lower knee represents configurations that maximise hydrogen permeation per reactor volume, where even marginal gains in volumetric productivity are almost impossible. Such knee points are well established in multi-objective optimisation as boundaries of optimal trade-off solutions, indicating where the balance between conflicting objectives shifts from beneficial to diminishing returns. They thus provide a rational basis for selecting practical reactor geometries depending on whether catalyst utilisation or reactor compactness is the primary design priority.

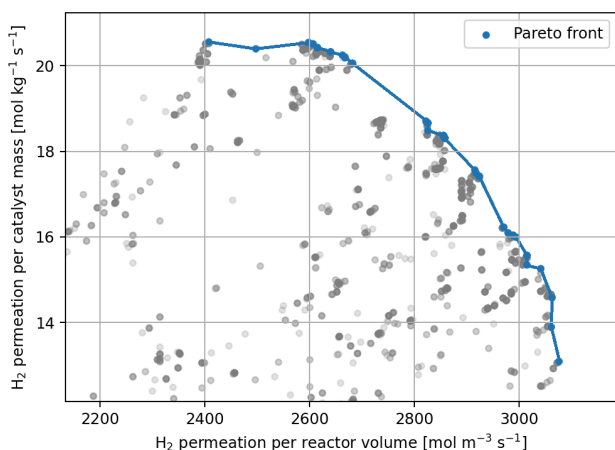


Fig. 5 Pareto front of the membrane reactor from NSGA-II.

Analysis of the Pareto-optimal reactor geometries shows a linear relationship between reactor length L and outer diameter d_{out} as shown in Fig. 6. It can be observed that among Pareto-optimal geometries, as length increases, diameter decreases accordingly. This dimensional shift reduces the total reactor volume, since the quadratic decrease in diameter outweighs the linear increase in length. At the same time, the catalyst mass increases as the annular catalyst region widens and the reactor becomes longer. Consequently, longer and slimmer reactors result in a higher hydrogen permeation per volume but a lower permeation per mass. This confirms that long and slim geometries are favoured when the volumetric objective is emphasised, and conversely short and wide designs are favoured when the catalyst mass objective is weighted more.

Following discussion of how prioritising one objective over the other affects the optimal reactor geometry, the impact of applying different weightings to the two objectives is further investigated. The results shown in Fig. 7 further illustrate how variation in the relative importance of the two objectives systematically shifts the optimiser towards different parts of the design space. Thus, giving more weight to catalyst-mass efficiency (red curve) favours

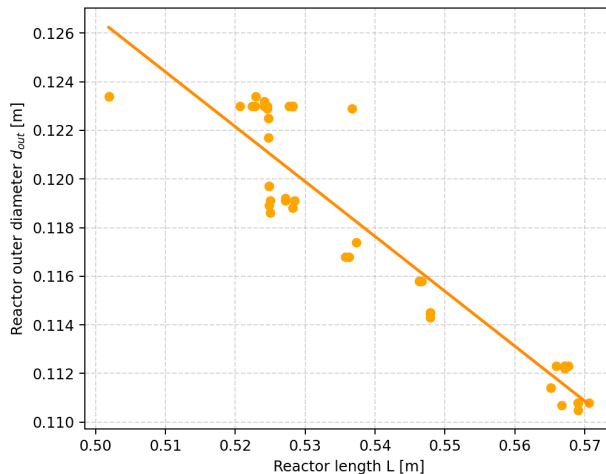


Fig. 6 Length and diameter of Pareto-optimal designs.

designs that minimise catalyst use, even if this involves larger reactor volumes, whereas weighting volumetric efficiency more heavily (green curve) drives the optimiser toward compact geometries that maximise hydrogen production per unit volume, at increased catalyst demand. Both of the optimisation converges under same optimiser, proving that NSGA-II has a high robustness for such optimisation problems. This further reinforces that the preferred reactor configuration depends on context. While both objectives are clearly relevant for transportation systems, their relative importance varies by sector. Aviation, for instance, is typically extremely volume constrained and may prefer volumetric productivity, while maritime applications—with much higher fuel throughput—may prioritise minimising catalyst consumption owing to its relatively high cost.

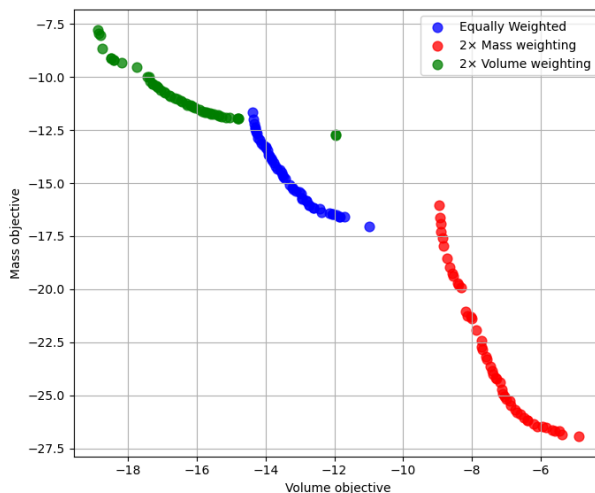


Fig. 7 Pareto fronts of objective functions at different weightings.

3.4 Reactor Sensitivity Analysis. Sensitivity analysis serves as a critical evaluation tool for assessing reactor robustness and verifying design adequacy across a range of operating conditions. By systematically varying key parameters, it enables a clear understanding of how strongly reactor performance depends on its operating environment and how resilient the system remains under off-design or extreme scenarios. In the present study, temperature, pressure, and ammonia inflow rate are selected as the principal variables influencing the ammonia conversion

profile. Examining their individual and combined impacts allows us to determine whether the reactor can consistently sustain high conversion efficiency and operational stability despite fluctuations or deviations from nominal conditions. This analysis ultimately supports the validation of the proposed reactor model and informs guidelines for safe and reliable operation.

The temperature-sensitivity analysis demonstrates in fig.8 shows its strong influence on ammonia conversion in the tubular reactor. As the temperature increases from 623 K to 773 K, the conversion profile exhibits a significant rise in both the rate of conversion and the overall conversion achieved at the reactor outlet. At lower temperatures (573–623 K), conversion remains negligible across the entire reactor length, indicating that the reaction rate is insufficient to drive meaningful cracking even over 0.6 m of catalyst bed. At moderate temperatures (673–693 K), conversion increases approximately linearly along the reactor and reaches 10–25% at the outlet. The most significant enhancement occurs at 723–773 K, where the reaction transitions into a kinetically favourable regime; ammonia conversion rapidly approaches 100% conversion. At 773 K in particular, conversion approaches completion over the optimal length. These results confirm that reactor performance is highly sensitive to temperature and that achieving high ammonia conversion requires operation within the upper thermal range of the design envelope. A similar trend is demonstrated in various past literature sources [15,18,19]. This insight is critical for establishing optimal operating conditions and ensuring sufficient hydrogen production even in the presence of temperature fluctuation during the operation.

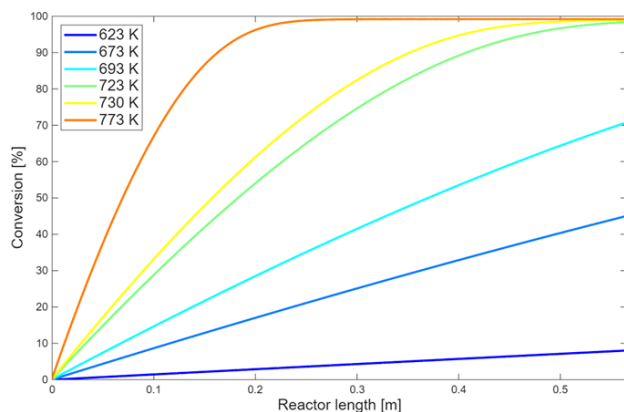


Fig. 8 Ammonia conversion sensitivity analysis with variation on operating temperature varying from 623-753 K.

The sensitivity analysis on ammonia inlet flow rate reveals a clear inverse relationship between feed rate and the rate of conversion along the reactor. Although all examined flow rates ultimately approach similar final conversion levels near the reactor outlet, the trajectory by which they reach this point differs substantially. Lower inlet flow rates (e.g. 5.0-7.5 mol/s) exhibit a markedly steeper conversion gradient, achieving more than 90% conversion within the first 0.2–0.3 m of reactor length. This indicates that, under reduced throughput, the residence time and local reaction conditions are highly favourable, enabling rapid ammonia cracking with minimal reactor volume. Conversely, as inlet flow increases, the conversion curve becomes progressively flatter. Higher flow rates (12.5–15.0 mol/s) require nearly the full reactor length to reach comparable conversion, reflecting the diminishing residence time and reduced kinetic driving force available under high-throughput operation.

This behaviour has important operational implications. While the reactor is capable of delivering high conversion across the full range of flow rates tested, the substantial reduction in conversion

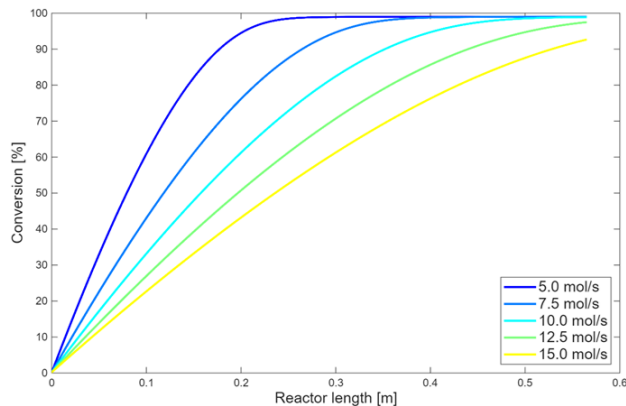


Fig. 9 Ammonia conversion sensitivity

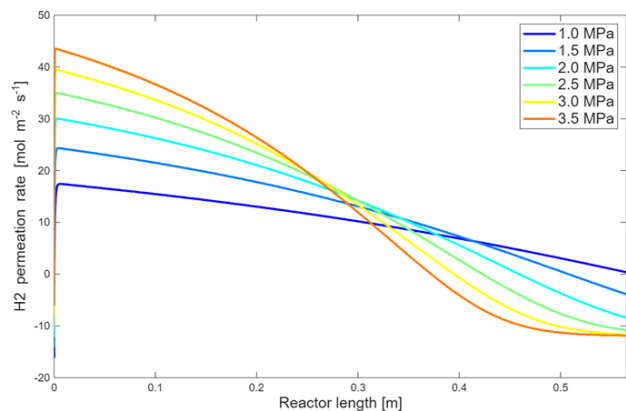


Fig. 10 Ammonia conversion sensitivity

rate at elevated inlet flow introduces potential control challenges during periods of extreme hydrogen demand. Under such conditions, the reactor may struggle to reach the required conversion level quickly enough, leading to temporary underperformance or the need for supplemental control strategies such as dynamic temperature adjustment, flow modulation, or buffer storage. The findings highlight the need for a robust control system capable of compensating for the sensitivity of the conversion dynamics to inlet flow rate.

4 Conclusions

This work has developed and applied a coupled reaction-separation model and a multi-objective optimisation framework for the design of a multi-tubular catalytic membrane ammonia cracker. An equilibrium-based reactor model is formulated that captures the essential features of ammonia thermal decomposition and predicts high ammonia conversions in good agreement with literature trends [15] and shows the strong equilibrium enhancement from continuous hydrogen removal through a palladium membrane. These results confirm that membrane-assisted ammonia cracking can achieve near-complete conversion under milder conditions than conventional thermal reactors, supporting its suitability for hydrogen generation from ammonia in transport applications.

The reactor model is then optimised using NSGA-II for two objectives – hydrogen permeation rate per unit reactor volume and rate per unit catalyst mass subjecting to five decision variables. The algorithm converged for the optimisation problem and moved to a smooth Pareto front, reporting a set of optimised decision variables. Additional analyses are conducted to evaluate the robustness of the optimisation under different objective weightings. The results proves that NSGA-II performs consistently and

reliably for this reactor design problem, showing strong potential for application to other complex reactor optimisation cases.

5 Outlook

Future work will focus on strengthening the applicability, and engineering robustness of the reactor model through both expanded simulation studies and system-level integration. In terms of reactor modelling, further simulations will investigate the influence of catalyst-bed voidage and radial temperature profiles on conversion behaviour and membrane performance. These analyses will support determining the reactor's operating envelope and provide validation against expected physical trends. Extending the current steady-state framework to a full dynamic model is a critical step, as a robust dynamic model enables transient behaviour, start-up/shutdown characteristics, and control-relevant sensitivities to be assessed. To complement one-dimensional modelling, a CFD simulation using COMSOL could be undertaken to evaluate local transport phenomena, non-uniformities, and potential design bottlenecks that reduced-order models cannot capture.

Beyond the reactor itself, the next phase of research should incorporate thermal management and process integration through a heat-exchanger network design. This could ensure that energy recovery, heat duties, and thermal coupling are accurately represented.

Insights from the sensitivity analysis also reveal several areas that require attention. First, safety and economic constraints were insufficiently represented in the current optimisation setup. Preliminary findings indicate that the selected membrane surface area may be excessively large, potentially resulting in triggering reverse permeation of hydrogen under certain conditions. Moreover, the pressure difference between the inner and outer tubes raises structural concerns regarding membrane tolerance. On the retentate side, the limited cross-sectional area of the outer tube may impose risks such as choked flow. Addressing these issues will require more robust NSGA-II constraint formulations that reflect realistic mechanical limits, operability considerations, and capital costs. Improving the formulation ensures future optimisation outputs are aligned with practical engineering feasibility.

6 Acknowledgement

The author gratefully acknowledges Siyu "Helen" Guo, PhD student at Imperial College London, for her guidance and assistance with the reaction modelling. ChatGPT is used for code revision, debugging support, and language refinement throughout the preparation of this work.

Nomenclature

Roman letters

- r_{NH_3} = Reaction rate [$\text{mol s}^{-1} \text{m}^{-3}$]
- p_i = Partial pressure of component i [Pa]
- k_{eq} = Equilibrium constant [-]
- M_i = Molar mass of component i [kg/mol]
- M_{mix} = Average molar mass of component i [kg/mol]
- x_i = Molar fraction of component i [-]
- T = Temperature [K]
- k = Reaction kinetic constant [-]
- $D_{k,m}$ = Knudsen diffusion coefficient of mixture [-]
- $D_{p,m}$ = Knudsen + binary diffusion coefficient of mixture [-]
- $D_{p0,m}$ = Intermediate variable for Knudsen diffusion calculation [-]
- P = reactor pressure [Pa]
- R = Universal gas constant, 8.314 [-]
- r_{av} = Correction parameter, $1e-6$ [-]
- h_T = Thiele modulus [-]

- Q_{H_2} = H_2 permeation rate of hydrogen [$\text{mol m}^{-2} \text{s}^{-1}$]
- r_p = catalyst particle diameter[m]
- $\dot{m}$ = mass flowrate [kg s]
- u_0 = Gas superficial velocity [m/s]
- A_{cs} = Cross sectional area of reactor unit m^2
- $\dot{n}_{H_2}$ = H_2 permeation rate [$\text{m}^2 \text{s}^{-1}$]
- V_{reactor} = Volume of single reactor unit [m^3]
- m_{cat} = Mass of catalyst load [kg]

Greek letters

- η = Catalyst efficiency factor [-]
- $\sigma_{mix}^2 \omega_{mix}$ = Acentric factor of mixture, an intermediate variable [-]
- ν = Intermediate variable relate to viscosity [-]
- μ_i = Viscosity of component i [Pa.s]
- μ_{mix} = Viscosity of the mixture [Pa.s]
- ϵ = Catalyst bed voidage, 0.7
- τ_{solid} = Catalyst tortuosity factor, 4
- ρ_{solid} = Catalyst bulk density [kg m^{-3}]
- ρ_p = Catalyst particle density [kg m^{-3}]
- ρ_{mix} = Density of gas mixture [kg m^{-3}]
- β = Exponential constant in rate law [-]
- Φ_{ij} = Molecular interaction parameter [-]

Superscripts and subscripts

- i = Chemical component, consist ammonia (NH_3), hydrogen (H_2) and nitrogen (N_2)
- mix = Gas mixture, corresponding to hydrogen, nitrogen and ammonia

References

- [1] "CO₂ Emissions – Global Energy Review 2025." URL <https://www.iea.org/reports/global-energy-review-2025/co2-emissions>.
- [2] Lutsey, Nicholas and Sperling, Daniel. "Greenhouse gas mitigation supply curve for the United States for transport versus other sectors." *Transportation Research Part D: Transport and Environment* Vol. 14 No. 3 (2009): pp. 222–229.
- [3] Wang, Ning, Huang, Shuai, Zhang, Zhifei, Li, Tie, Yi, Ping, Wu, Dawei and Chen, Gen. "Laminar burning characteristics of ammonia/hydrogen/air mixtures with laser ignition." *International Journal of Hydrogen Energy* Vol. 46 No. 62 (2021): pp. 31879–31893.
- [4] Di Carlo, Andrea, Dell'Era, Alessandro and Del Prete, Zaccaria. "3D simulation of hydrogen production by ammonia decomposition in a catalytic membrane reactor." *International journal of hydrogen energy* Vol. 36 No. 18 (2011): pp. 11815–11824.
- [5] Moudio, Natacha Diane Ngasse, Bian, Xiao-Qiang and Chinamo, Dennis Sabato. "Liquid hydrogen carriers for clean energy systems: A critical review of chemical hydrogen storage strategies." *Fuel* Vol. 404 (2026): p. 136329.
- [6] Valera-Medina, Agustin, Xiao, Hua, Owen-Jones, Martin, David, William IF and Bowen, Philip John. "Ammonia for power." *Progress in Energy and combustion science* Vol. 69 (2018): pp. 63–102.
- [7] Cerone, Nadia, Zito, Giuseppe Domenico, Florio, Carmine, Fabbiano, Laura and Zimbardi, Francesco. "Recent advancements in Pd-based membranes for hydrogen separation." *Energies* Vol. 17 No. 16 (2024): p. 4095.
- [8] Conde, Julio J, Maroño, Marta and Sánchez-Hervás, José María. "Pd-based membranes for hydrogen separation: review of alloying elements and their influence on membrane properties." *Separation & Purification Reviews* Vol. 46 No. 2 (2017): pp. 152–177.
- [9] Maccarrone, Domenico, Giorgianni, Gianfranco, Italiano, Cristina, Perathoner, Siglinda, Centi, Gabriele and Abate, Salvatore. "Comparative analysis of hydrogen production from ammonia decomposition in membrane and packed bed reactors using diluted NH₃ streams." *International Journal of Hydrogen Energy* Vol. 82 (2024): pp. 513–522.
- [10] Zhang, Xiaochen, Li, Chaoqun, He, Zhenyu and Han, Te. "Integration of ammonia synthesis gas production and N₂O decomposition into a membrane reactor." *Industrial & Engineering Chemistry Research* Vol. 60 No. 7 (2020): pp. 3066–3072.
- [11] Cecchetto, Valentina, Di Felice, Luca and Gallucci, Fausto. "Advances and Perspectives of H₂ production from NH₃ decomposition in membrane reactors." *Energy & Fuels* Vol. 37 No. 15 (2023): pp. 10775–10798.
- [12] Deb, Kalyanmoy, Pratap, Amrit, Agarwal, Sameer and Meyarivan, TAMT. "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE transactions on evolutionary computation* Vol. 6 No. 2 (2002): pp. 182–197.
- [13] Konak, Abdullah, Coit, David W and Smith, Alice E. "Multi-objective optimization using genetic algorithms: A tutorial." *Reliability engineering & system safety* Vol. 91 No. 9 (2006): pp. 992–1007.
- [14] Coello, CA Coello. "Evolutionary multi-objective optimization: a historical view of the field." *IEEE computational intelligence magazine* Vol. 1 No. 1 (2006): pp. 28–36.
- [15] Guo, Siyu and Mercangöz, Mehmet. "Design and Evaluation of a Multitubular Catalytic Membrane Reactor for Onboard Ammonia Cracking in Commercial Aviation." *ASME Open Journal of Engineering* Vol. 4 (2025).
- [16] Tang, Xin-Yuan, Yang, Wei-Wei, Ma, Xu and Cao, Xiangkun Elvis. "An integrated modeling method for membrane reactors and optimization study of operating conditions." *Energy* Vol. 268 (2023): p. 126730.
- [17] Collins, John P, Way, J Douglas and Kraisuwansarn, Nichakorn. "A mathematical model of a catalytic membrane reactor for the decomposition of NH₃." *Journal of membrane science* Vol. 77 No. 2-3 (1993): pp. 265–282.
- [18] Sima, Dewen, Wu, Haojin, Tian, Koukou, Xie, Shijun, Foo, Joel Jie, Li, Shuirong, Wang, Duo, Ye, Yueyuan, Zheng, Zhifeng and Liu, Yun-Quan. "Enhanced low temperature catalytic activity of Ni/Al–CeO₂ for hydrogen production from ammonia decomposition." *International Journal of Hydrogen Energy* Vol. 45 No. 16 (2020): pp. 9342–9352.
- [19] Mukherjee, Shreya, Devaguptapu, Surya V, Sviripa, Anna, Lund, Carl RF and Wu, Gang. "Low-temperature ammonia decomposition catalysts for hydrogen generation." *Applied Catalysis B: Environmental* Vol. 226 (2018): pp. 162–181.

Computational modelling of thrombolysis: Effect of varying therapeutic drug and characteristics of the blood clot

Weber Quiterie and Dandjee Nora

Department of Chemical Engineering, Imperial College London, U.K.

ABSTRACT Blood clots form in blood vessels and can cause serious health issues, such as strokes. Their treatment relies on thrombolytics, such as alteplase (ATP), the most common, and tenecteplase (TNK), recently approved. Previous computational studies on arterial thrombi usually involve multiple design assumptions and disregard multiple blood clot properties variations. This research aims to bring new insights on thrombi through computational modelling, by investigating how thrombolysis within the middle cerebral artery and pulmonary veins is influenced by varying the therapeutic drug, blood clot type, position and length, using a PKPD model. Results showed that TNK exhibited longer overall lysis times, but dissolved fibrin-dense regions more effectively while displaying more consistent lysis behaviour due to its strong fibrin specificity. Clot composition strongly impacted thrombolysis as “red” arterial thrombi lysed fastest due to their moderate voidage and higher permeability, and inversely for “white” arterial thrombi. Venous thrombi limited the drug penetration due to unfavourable venous haemodynamics. Clots located downstream exhibited delayed lysis, especially for ATP, due to pre-clot reactions during drug transport. Impact of thrombus’s length depended on clot’s location where longer clots were lysed faster at the inlet, while the opposite effect was observed for deeper locations. Overall, the study highlights the complex interactions between clot characteristic, drug properties and flow conditions on thrombolytic efficiency.

KEYWORDS thrombolysis modelling, fibrin specificity, alteplase, tenecteplase, clot heterogeneity, venous thrombus

1. INTRODUCTION

Thrombi, or blood clots, are a complex mixture of red blood cells (RBCs), fibrin, platelets and other cellular components that form within arteries or veins (Diamond, 1999). While clot formation is a natural protective response to vascular injury, excessive coagulation or thrombosis can obstruct blood vessels and restrict blood supply to vital organs such as the heart or lungs (Banerjee, Chisti & Banerjee, 2004). Untreated, thrombosis represents a major public health challenge, contributing to one in four deaths each year. It is the cause of some of the most widespread cardiovascular conditions, including ischemic stroke when clots lodge in brain arteries and venous thromboembolism for clots formed in venous circulations. Despite its consequences, public awareness and scientific understanding of thrombosis remain limited. A deeper understanding of the complexities and factors influencing thrombolysis is crucial to help find solutions for cardiovascular diseases (World Health Organization, 2025).

Current treatment for acute thrombotic diseases mainly relies on pharmacological thrombolysis, in which fibrinolytic agents such as tissue plasminogen activator (tPA) are administered to dissolve the fibrin clots via a series of complex biochemical reactions. Alteplase (ATP), a recombinant tPA, is thus far the most well-established and studied agent due to its lysis efficiency and its moderate fibrin specificity (Adivitiya & Khasa, 2017). However, its clinical utility is limited by a short half-life of 6 minutes, susceptibility to plasminogen activator inhibitor-1 (PAI-1), and an increased risk of intracranial haemorrhage (Peters & Paciullo, 2015) (Gu et al., 2019a) (Adivitiya & Khasa, 2017). Tenecteplase (TNK), a genetically engineered tPA variant, has recently attracted attention as a potential alternative due to its longer half-life and greater fibrin specificity, which allow for a bolus injection. (Yang, Gu & Xu, 2023).

However, despite TNK’s recent U.S. Food and Drug Administration (FDA) approval, comparative data

between alteplase and tenecteplase remain limited, with few studies directly evaluating their relative thrombolytic performance (touchNEUROLOGY, 2025) (Menon et al., 2022). This gap exists because drug lysis efficiency varies across cases, as evidence indicates that treatment success is influenced not only by the drug’s kinetic properties but also by the underlying characteristics of the clot itself. Factors such as clot composition, internal structure, and anatomical location can significantly affect thrombus response to fibrinolytic agents (Jolugbo & Ariëns, 2021) (Narwal, Whyte & Mutch, 2023).

Assessing the efficiency of thrombolytic therapy under various conditions is challenging in both in vitro and in vivo (Nikitin et al., 2021). Consequently, mathematical and computational modelling enables controlled drug transport, biochemical reactions, and clot degradation under physiologically relevant conditions (Bannish & Hudson, 2021).

This report aims to address these limitations by systematically comparing the thrombolytic efficiency of ATP and TNK across a range of physiologically realistic scenarios using a thrombolysis computational model. This include comparing arterial and venous environments, varying clot location within the vessel, and thrombus size. The objective is to get more insights into how thrombolytic therapy is influenced by drug properties and clot characteristics, with the goal of designing more effective treatment strategies.

2. BACKGROUND

Arterial thrombi are among the most extensively studied form of blood clots, largely because of their central role in acute cardiovascular diseases (Mereuta et al., 2021). Di Meglio et al. (2019) demonstrated that thrombi in the middle cerebral artery (MCA) often exhibit a heterogeneous structure, characterised by a dense, platelet-rich shell and a loose, RBC-rich core. They are typically classified as “red” and “white” thrombi (Jolugbo & Ariëns, 2021). “Red” blood clots are characterised by their higher proportion of RBCs

entrapped in fibrin. On the other hand, “white” blood clots are smaller and are primarily composed of platelets. This type is also less common, representing only 15% of arterial obstruction cases in cerebral arteries. Fibrin content is similar across both types of clots (Mereuta et al., 2021).

Additionally, Gu et al. (2019b) demonstrated that clot properties can vary by arterial location, with clots in the M1 and M2 segments of MCA often differing in size and internal structure. These can significantly influence both drug delivery and the effectiveness of thrombolytic treatment.

In contrast to arterial thrombi, venous clots have been less extensively studied, and available evidence doesn't support the presence of a heterogeneous structure (Chernysh et al., 2020). This suggests that venous thrombi have a more uniform internal structure, primarily composed of fibrin and RBCs, with no defined core-shell structure as observed in arterial clots (Narwal, Whyte & Mutch, 2023). This homogeneous structure implies different drug transport behaviour and thus different thrombolytic response (Chernysh et al., 2020).

Numerous computational models were developed to understand the mechanisms governing thrombolysis. For instance, Piebalgs et al. (2018) investigated the influence of vessel geometry and clot size, while Yang, Gu and Xu (2023) explored the impact of the fibrinolytic agent used. Although these studies provided valuable insights, most oversimplify the process by treating the clot as a uniform fibrin mass, ultimately failing to capture the influence of clot heterogeneity on thrombolysis. Furthermore, venous thrombi are modelled even less frequently, despite their clinical relevance.

This report addresses the current knowledge gap by comparing the thrombolytic performance of ATP and TNK across a range of clinically relevant scenarios, incorporating a model that considers the heterogenous structure of arterial clots (Narongchartsopon et al., 2025).

3. METHODS

3.1. Computational model

The dissolution of a blood clot under thrombolytic therapy involves numerous biochemical reactions between tPA, the fibrinolytic proteins and the fibrin within the thrombus. The process was simulated using a computational model for thrombolysis developed by Gu et al. (2019b). It includes a two-compartment pharmacokinetics and pharmacodynamics (PKPD) model and considers the blood flow influence on the system.

The PKPD model describes the co-influence between the tPA and the body after drug injection. This involves the equations related to drug diffusion, and reactions between the fibrinolytic proteins and their production and elimination by the body. Figure 3.1.1 represents the reactions and transfers considered in the PKPD model, with their kinetics detailed in section 9.4 in the Supplementary Information.

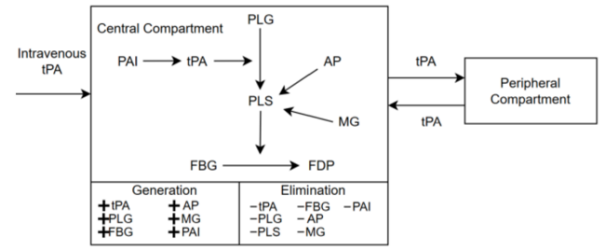


Figure 3.1.1: Two-compartment systemic PKPD model (Narongchartsopon et al., 2025)

The following equations describe the transfer of tPA between the central and peripheral compartments (Equations 3.1.1 and 3.1.2), representing the plasma and the rest of the body, and the change in concentration for each component (Equation 3.1.3):

$$\frac{dC_{tPA,c}}{dt} = \left[\frac{I}{M_{W,tPA} V_{plasma}} + k_{p \rightarrow c} C_{tPA,p} \right] - [(k_{el,tPA} + k_{c \rightarrow p}) C_{tPA,c}] + [G_{tPA}] + [R_{tPA}^{plasma}] \quad (3.1.1)$$

$$\frac{dC_{tPA,p}}{dt} = [k_{c \rightarrow p} C_{tPA,c}] - [k_{p \rightarrow c} C_{tPA,p}] + [0] + [0] \quad (3.1.2)$$

$$\frac{dC_i}{dt} = [0] - [k_{el,i} C_i] + [G_i] + [R_i^{plasma}] \quad (3.1.3)$$

Where $i = \{PLG, PLS, FBG, AP, MG, PAI, PLS-AP\}$, $k_{p \rightarrow c}$ and $k_{c \rightarrow p}$ represents the kinetic constants of tPA which travels between the central and peripheral compartments, and I is the tPA infusion rate.

The blood flow characteristics in the obstructed blood vessel are defined through convection and fluid mechanics equations. First, the bifurcation after the internal carotid artery (ICA) is considered, as it divides into the MCA and the anterior cerebral artery (ACA), with the MCA being the obstructed artery. The complex geometry of blood vessels is simplified to a straight and circular cylinder. This is to ease the simulation while keeping pressure-flow characteristics realistic. The blood flow velocity U is calculated using Darcy's law:

$$U = \frac{1}{\varepsilon_{clot}} \frac{k_t}{\mu} \Delta P_x \quad (3.1.4)$$

This equation depends on the pressure drop across the clot length ΔP_x , permeability k_t (calculated through Davies' equation and Either's model), blood viscosity μ and clot porosity ε_{clot} , which vary with the blood clot composition and degradation state (see section 9.9 in the Supplementary Information.). The pressure gradient is assumed to be constant during the blood clot lysis. In reality, as the clot gets dissolved, the pressure drop across the thrombus increases, allowing further blood flow in the obstructed region, which heavily impacts the process (Piebalgs & Xu, 2015). Hence, Equation 3.1.4 yields the blood flow rate along the MCA throughout the simulation duration. The arterial geometry is presented in Figure 3.1.2.

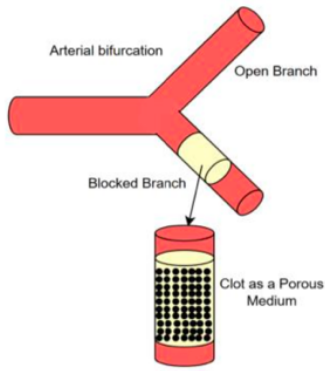


Figure 3.1.2: Schematic representation of the arterial bifurcation, with the obstructed artery (Narongchartsopon et al.)

The obstructed blood vessel is divided into two phases: the free phase, representing the flowing plasma, and the bound phase, representing the fibrin in the blood clot. The pharmacodynamics in the obstructed blood vessel were modelled using the same reaction network as presented in Figure 3.1.1, whilst also accounting for the fibrinolytic reactions occurring between and within free and bound phases, represented in Figure 3.1.3. Their kinetics are detailed in sections 9.4 and 9.5 in the Supplementary Information. The tPA transfer between the central and peripheral compartments is not considered for this section. The components' concentrations are calculated using the following equations, for the free and bound phases, respectively:

$$\frac{\partial C_i}{\partial t} = -\frac{\partial(UC_i)}{\partial x} + D_i \frac{\partial^2 C_i}{\partial x^2} + R_i^{\text{tot}} \quad (3.1.5)$$

$$\frac{\partial n_i}{\partial t} = R_i^{\text{clot}} \quad (3.1.6)$$

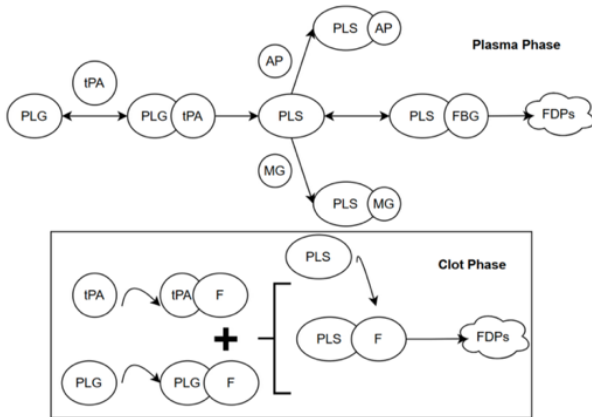


Figure 3.1.3: Reaction mechanisms between the free and bound phases

This model was then improved by accounting for the clot heterogeneity (Narongchartsopon et al., 2025). A new blood clot structuration is introduced, dividing the thrombus into two areas: an RBC-rich core and two fibrin- and platelet-rich shells, located at its front and rear (see Figure 3.1.4). Each shell represents a hundredth of the total clot length. This improvement enables more realistic simulations, yielding more accurate results, and allows further investigations into the thrombus's structure. This also involves new considerations, such as the RBCs release, calculated using the hindrance term (Narongchartsopon et al., 2025):

3

$$H(x,t) = \begin{cases} 1 - \frac{I}{-m \left(E_L - \frac{E_{L,crit}}{k} \right)}, & E_L < E_{L,crit} \\ 0, & E_L \geq E_{L,crit} \end{cases} \quad (3.1.7)$$

Where E_L represents the extent of lysis calculated by:

$$E_L(x,t) = 1 - \frac{n_{tot}}{n_{tot,0}} \quad (3.1.8)$$

With n_{tot} being the total number of binding sites. Each layer of the blood clot has a distinct permeability, fibrin density and reaction parameters, determined at each time step. Details of the equations solved are found in section 9.9 in the Supplementary Information.

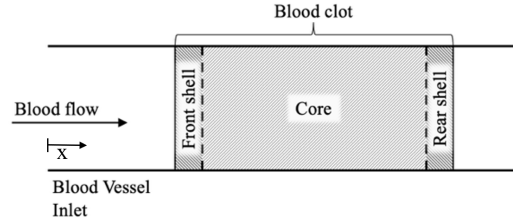


Figure 3.1.4: Structure of the modelled blood clot

From the plasma inlet conditions and blood clot initial properties, the computational model solves the set of ODEs in MATLAB to calculate the concentration of the components in both phases at each time and spatial step.

The model includes multiple design assumptions, which impact the accuracy of the results. The blood vessel geometry is simplified to a straight and circular cylinder, and the system is set as only one-dimensional to simplify the simulation process. However, this simplification models the blood flow as fully laminar, when it usually becomes turbulent at high velocity or at blood vessel bifurcation (Sidik & Mazumdar, 1994). Furthermore, it implies a 100% obstruction from the blood clot, assuming an identical dissolution behaviour for each point across the vessel's cross-sectional area. Thrombi usually only partially block the blood flow (Chen et al., 2015) and therefore also have a top shell, which involves variations in the thrombolytic action and response.

The model is set to end when the extent of lysis exceeds 95% ($E_{L,crit}$) at every layer, considered as complete lysis, or if the blood clot hasn't dissolved in 500 minutes, considered as incomplete lysis. The properties calculated and returned by the model include the concentration of each protein in the free and bound phases, along with the lysis time and the extent of lysis and resistance profile in the system. See section 9.8 in the Supplementary Information for the model specifications.

3.2. Parameters varied

The impact of multiple parameters on the dissolution of blood clots has rarely been studied in previous works, which often ignore important properties in their model. The blood clots are usually assumed to have a unique fibrin, RBC, and platelet composition and their heterogenous structure isn't considered, although playing a major role in blood clot dissolution (Jolugbo & Ariëns, 2021). In this study, the administration of different drugs was simulated, and multiple blood clot properties were varied to observe their impact on thrombolysis.

The first parameter is the choice of therapeutic drug injected. They act as a thrombolytic by activating plasminogen to form plasmin, as described earlier by the reactions. This weakens the fibrin in the blood clot, ultimately breaking it down completely. The first drug considered is ATP, which is the most administered for blood clot cases. It has a high fibrin affinity due to the presence of tPA, but a short half-life of around 6 minutes, which indicates a quick action, but a rapid dissipation of its effects as well (Adivitiya & Khasa, 2017). For safety and performance reasons, ATP has a high dose of 0.9mg/kg. It is administered through two injections: 10% of the required volume is injected in the first minute of treatment, while the remaining 90% is infused throughout the following 60 minutes. The second drug analysed is TNK, which was only recently approved; therefore, there is limited studies on its performance on blood clot lysis. It is a modified version of ATP, with a greater PAI-I resistance and a longer half-life, ranging between 10 and 24 minutes, depending on its mutations (Adivitiya & Khasa, 2017). Following the current standard, it is administered through a single injection lasting 5 seconds, with a dose of 0.25mg/kg. The dosage regimen and parameters for both drugs are listed in section 9.2 in the Supplementary Information. The drug choice also affects some reaction kinetics, detailed in sections 9.4 and 9.5 in the Supplementary Information.

The second parameter analysed in this study is the blood clot type. Two main types are extracted from MCA segments: “red” and “white” blood clots (Mereuta et al., 2021). They notably differ in RBC and platelet content in their core and shell, which has a major influence on their lysis. In this study, RBCs are neglected in the shell, and the same applies for platelets in the core, as they account for only a very small volume fraction in each area (Mereuta et al., 2021). Blood clots located in the pulmonary veins will also be modelled to compare their lysis with arterial thrombi. No shell-core structure has yet been proved to exist for venous blood clots. Compared to arterial thrombi, they have a high fibrin content and very low voidage in real-life situations. Their composition was heavily simplified to obtain sensible results in the simulations: they are assumed to only contain bundled fibrins, ignoring the other components’ presence, such as RBCs and platelets (Chernysh et al., 2020). The difference in blood vessels has multiple impacts on the system, notably on its blood flow due to their pressure and size differences (Widrich & Shetty, 2024). The compositions of both arterial blood clots and venous thrombi are detailed in Table 3.2.1 and Table 3.2.2.

Table 3.2.1: Composition of “red” and “white” blood clots

Thrombus type	RBC in the core	Fibrin in the core	Platelet in the shell	Fibrin in the shell
“Red” blood clot	46%	1%	20%	30%
“White” blood clot	6%	1%	55%	30%

Table 3.2.2: Composition of venous blood clots

Thrombus type	Fibrin fraction	Voidage
Venous blood clot	17.9%	82.1%

The third parameter considered in this study is the position of the blood clot in the blood vessel. Blood clots obstructing the M1 and M2 segments of the MCA were considered, modelled respectively 0.1 mm and 7.34 mm away from the vessel inlet. This position variation has an impact on the time taken by the drug to reach the blood clot. Furthermore, blood clots extracted from the M2 segment are usually shorter (Rossi et al., 2021), which was also respected. The venous blood clots were located at the same distances in the pulmonary vein to compare all types of clots more easily.

The last blood clot property varied was its length. Multiple studies found a correlation between the thrombus size and the extracted blood clot type (Mereuta et al., 2021), as well as the blood clot position in the MCA (Rohan et al., 2014). As previously mentioned, “red” and M1-segment clots are, on average, longer than “white” clots and M2-segment clots respectively. However, no study considers both properties when analysing the extracted blood clots. Therefore, the clot lengths simulated were determined from the values obtained by Mereuta et al. (2021); the average and greater length values were attributed to the “red” type, and the smaller and average ones to the “white” type, each time considering their arterial position. Venous blood clot took both the “red” and “white” lengths to obtain comparable results, again respecting the thrombus position. The lengths simulated, depending on the blood clot type and position, are shown in Table 3.2.3.

Table 3.2.3: Simulated lengths of “red” and “white” blood clots

Thrombus type	Inlet (M1 segment)	7.34mm (M2 segment)
“Red” blood clot	11 mm, 15 mm	6 mm, 8 mm
“White” blood clot	8 mm, 11 mm	4.5 mm, 6 mm
Venous blood clot	8 mm, 11 mm, 15 mm	4.5 mm, 6 mm, 8 mm

The computational model for thrombolysis was used to run the simulations for each combination of those four parameters. Tables 3.2.4 and 3.2.5 list all thrombus scenarios simulated in this study, along with their corresponding parameters.

Table 3.2.4: Simulation details, listing the blood clot type (including blood vessel), position and length modelling ATP injection

#	Drug injected	Thrombus type	Thrombus position	Thrombus length
A01	ATP	Venous	0.1 mm	8 mm
A02				11 mm
A03				15 mm
A04			7.34 mm	4.5 mm
A05				6 mm
A06				8 mm
A07		MCA: “white” blood clot	0.1 mm	8 mm
A08				11 mm
A09			7.34 mm	4.5 mm
A10				6 mm
A11		MCA: “red” blood clot	0.1 mm	11 mm
A12				15 mm
A13			7.34 mm	6 mm
A14				8 mm

Table 3.2.5: Simulation details, listing the blood clot type (including blood vessel), position and length modelling TNK injection

#	Drug injected	Thrombus type	Thrombus position	Thrombus length
T01	TNK	Venous	0.1 mm	8 mm
T02				11 mm
T03				15 mm
T04			7.34 mm	4.5 mm
T05				6 mm
T06				8 mm
T07		MCA: “white” blood clot	0.1 mm	8 mm
T08			7.34 mm	4.5 mm
T09			7.34 mm	6 mm
T10			7.34 mm	6 mm
T11		MCA: “red” blood clot	0.1 mm	11 mm
T12			7.34 mm	15 mm
T13			7.34 mm	6 mm
T14			7.34 mm	8 mm

4. RESULTS

All the simulations were carried out using the computational model, and the lysis time and extent-of-lysis graphs are listed in Table 9.10.1 in the Supplementary Information.

Note that the lysis time for thrombi in the pulmonary veins, being dissolved by TNK (simulations T01 to T06), exceeded the 500-minutes lysis time limit. Also note that the simulations of the “white” blood clot of length 4.5mm (simulations A09 and T09) did not model its heterogeneous structure, and the results were considered inaccurate. Therefore, these 8 simulations were not analysed when comparing results.

4.1. Therapeutic Drug Injected

As mentioned previously, the therapeutic drug injected influences multiple aspects of the blood clot lysis, due to their differences in half-life, resulting in different dosage regimens, and differences in the kinetic parameters, varying the fibrin interactions.

The lysis times obtained from all simulations are represented in Figure 4.1.1. The simulations A01 to A06, A09 and T09 were not analysed as the lysis was considered incomplete or the results inaccurate.

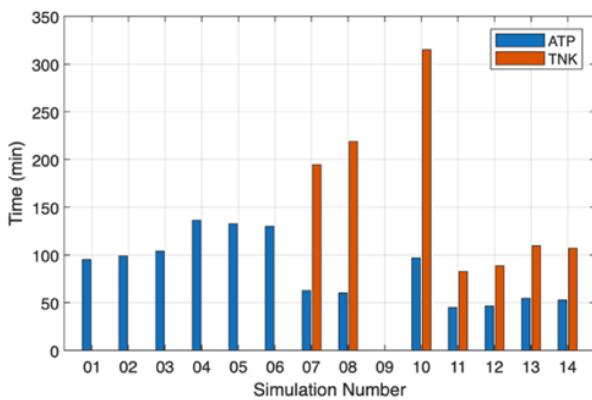


Figure 4.1.1: Completed lysis time for all simulations

This graph clearly indicates a significant increase in lysis time when injecting TNK instead of ATP, independently of the clot’s type, position or length. For the same blood clot parameters, TNK takes on average 2.5 times longer than ATP to complete the lysis.

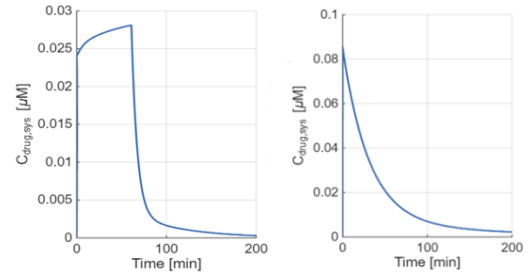


Figure 4.1.2: Drug concentration in the body over time; (a) ATP on the left, (b) TNK on the right

Figures 4.1.2 (a) and (b) plot the drug concentrations in the system against time from the onset of injection. As expected, the two concentration profiles mirror their drug’s injection regime: ATP’s concentration increases logarithmically, reaching its maximum of $0.028\mu\text{M}$ an hour after the injection starts, and drastically decreases exponentially afterwards. TNK peaks at $0.085\mu\text{M}$ in the first few instants after injection, before decreasing exponentially.

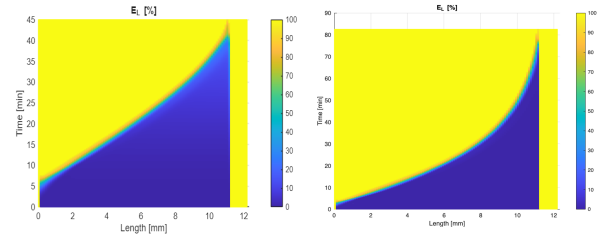


Figure 4.1.3: Extent of lysis throughout time against the blood clot length; (a) A11 on the left, (b) T11 on the right

As shown in Figures 4.1.3 (a) and (b), the extent-of-lysis trends vary between ATP and TNK. ATP’s profile tends to be more linear to logarithmic, depending on the blood clot scenario, whereas TNK exhibits a more exponential profile.

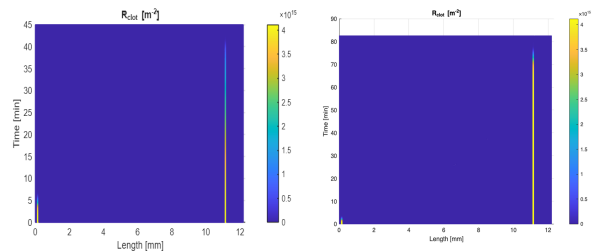


Figure 4.1.4: Resistance throughout time against the blood clot length; (a) A11 on the left, (b) T11 on the right

Figures 4.1.4 (a) and (b) show the resistance in the clot throughout its dissolution. In the ATP cases, the dissolution of the rear shell starts early in the process but takes almost the whole simulation time to be dissolved. For TNK, the rear shell starts dissolving once the rest of the thrombus has been lysed, and its lysis takes only a few minutes.

4.2. Blood Clot Type

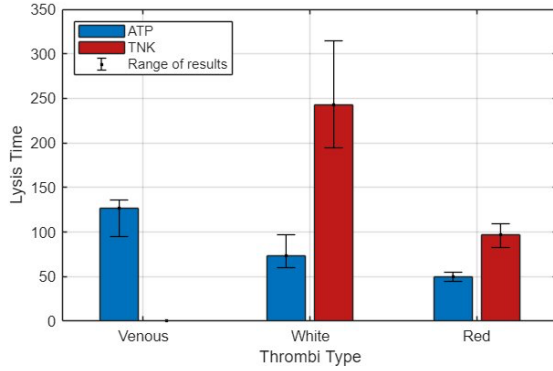


Figure 4.2.1: Average completed lysis time across venous, white and red thrombi, for ATP and TNK (Black error bars represent the full range of lysis time results)

TNK did not achieve full venous clot dissolution within the defined simulation limit of 500 minutes. These results were therefore excluded as physiologically irrelevant. Figure 4.2.1 summarises the complete lysis times for “red” arterial, “white” arterial, and venous thrombi treated with both ATP and TNK. Across both drugs, “red” clots consistently got lysed the fastest, whereas venous thrombi were the most resistant, requiring on average 50- and 125-minute dissolution time respectively for each type, administering ATP.

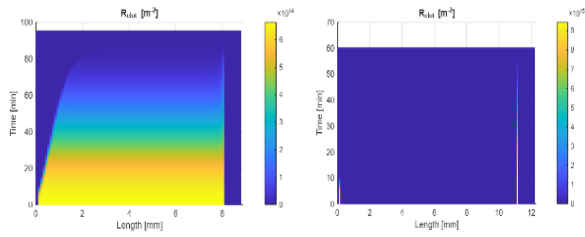


Figure 4.2.2: Resistance throughout time against the blood clot length; (a) A2 on the left, (b) A8 on the right

Despite being modelled as homogeneous structures, Figure 4.2.2 (a) shows that venous clots still experienced non-localised resistance. In contrast, the resistance exhibited by arterial clots are majorly at both shells’ locations, represented by two vertical lines (see Figure 4.2.2 (b)) (Di Meglio et al., 2019).

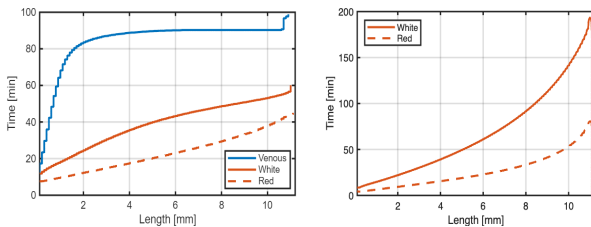


Figure 4.2.3: Contour of 100% extent of lysis throughout time against the blood clot length, across blood clot type evaluated; (a) ATP on the left, (b) TNK on the right

Figures 4.2.3 (a) and (b) demonstrate that the extent-of-lysis pattern differed between the two drugs. While the internal clot structure strongly influenced ATP’s behaviour, TNK’s lysis pattern remains similar across all scenarios. For instance, both “red” and “white” extent-of-lysis profiles exhibit an exponential-like trend: the first part of the clot was degraded rather quickly along its length before slowing down around 8 mm.

Figure 4.2.3 (a) shows that venous clots took the longest to be penetrated by ATP. However, once it does, the remaining clot dissolved uniformly, as suggested by the plateau in the extent-of-lysis curve. In arterial clots, ATP degraded the outer region of the “red” thrombi faster than that of “white” thrombi, followed by an exponential-like lysis trend for the rest of the clot, similar to TNK’s profile. “White” thrombi, on the other hand, exhibited an overall slower lysis rate across its length (Figures 4.2.3 (a) and (b)).

4.3. Blood Clot Position

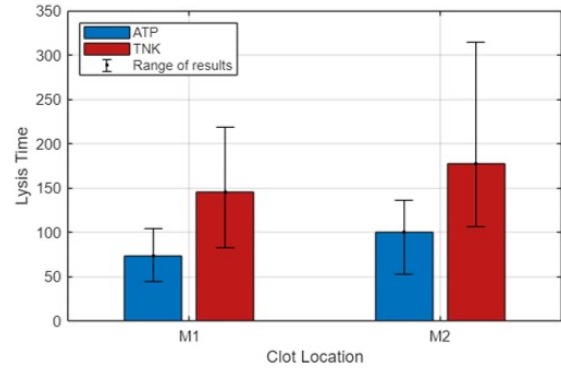


Figure 4.3.1: Average completed lysis time across M1 and M2 positioned thrombi, for ATP and TNK (Black error bars represent the full range of lysis time results)

Figure 4.3.1 summarises the average lysis time for both fibrinolytic agents across M1 and M2-positioned clots. It can be concluded that clots located further downstream in the MCA required an average of 25 more minutes to achieve complete lysis compared to those located in M1.

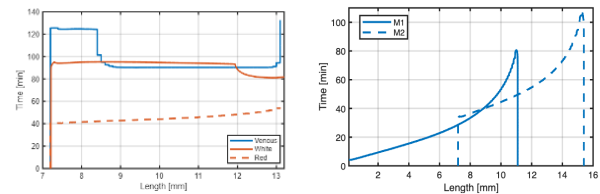


Figure 4.3.2: Contour of 100% extent of lysis throughout time against the blood clot length, for M2 location across blood clot type evaluated; (a) ATP on the left, (b) TNK on the right

Figure 4.3.2 (a) shows that ATP exhibited a complex and non-consistent behaviour across scenarios. In particular for both “white” and venous thrombi, a counterintuitive phenomenon is observed in which the core and distal shell were lysed earlier than the front shell. Additionally, the model also predicted lysis in regions that the agent had not yet exposed. Furthermore, this pattern was not observed in “red” thrombi. The latter exhibited lysis similar to those of the M1 segment, with only an extra time delay.

On the other hand, once TNK reached the clot surface, the lysis profile pattern remained almost identical between M1 and M2, indicating that TNK lysis dynamics were largely insensitive to clot position (see Figure 4.3.2 (b)).

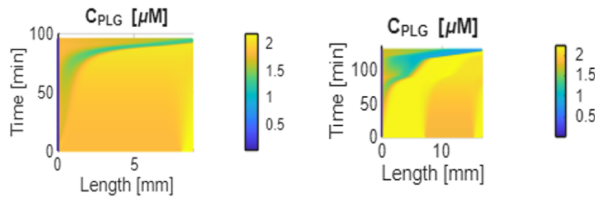


Figure 4.3.33: Free-phase plasminogen concentration throughout time against the blood clot length; (a) A1 on the left, (b) A6 on the right

Finally, a change in free-phase plasminogen concentration was observed for most M2-positioned clots simulations. For instance, Figures 4.3.33 (a) and (b) shows that the concentration of plasminogen slightly reduced ahead of the clot in the M2-clots simulations as seen by the green area ahead of the clot being more pronounced in Figure 4.3.3 (b) than (a).

4.4. Blood Clot Length

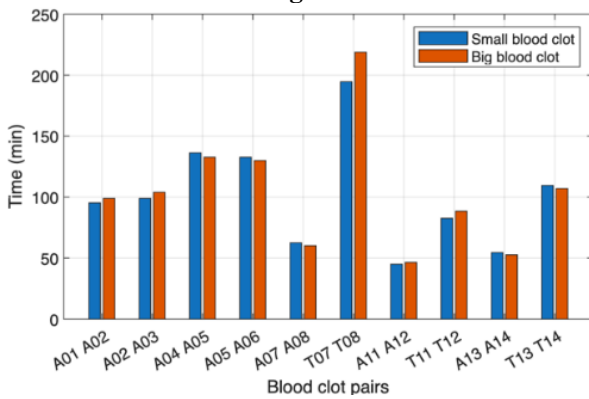


Figure 4.4.1: Completed lysis time comparison for pairs of blood clots in which length is the only varying parameter

The simulations were paired such that each pair contains two blood clots with only a difference in length. The lysis times of each pair are plotted in Figure 4.4.1, which shows only a slight time variation in each pair. The time ratio of each pair $r_{tPA,j}/t_{PA,i}$ was calculated to better compare these results:

$$r_{tPA,j}/t_{PA,i} = t_{PA,j}/t_{PA,i} \quad (4.4.1)$$

With t the lysis time, t_{PA} the drug injected (A for ATP and T for TNK), i and j the simulation number (with j the longer blood clot).

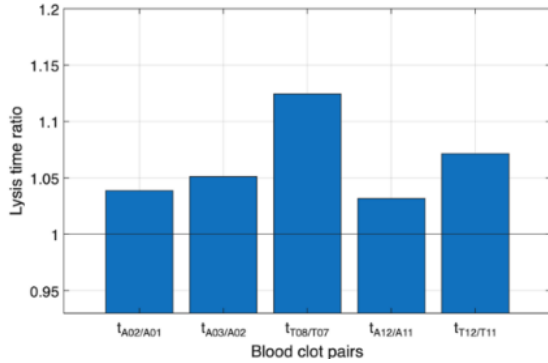


Figure 4.4.2: Blood clot pairs time ratios greater than 1

Figure 4.4.2 gathers all the blood clot pairs with a time ratio greater than 1, indicating greater lysis time for longer blood clots. All these blood clots are located at the

inlet of the obstructed blood vessel. Their ratios vary between 1.032 and 1.124.

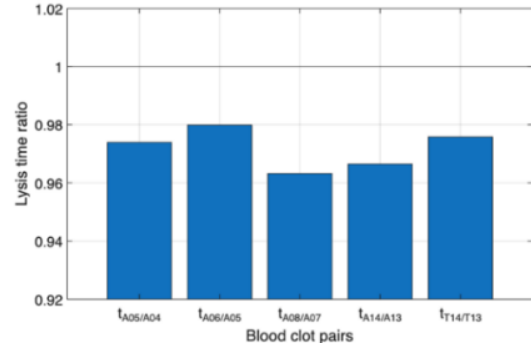


Figure 4.4.3: Blood clot pairs time ratios lower than 1

Figure 4.4.3 plots the blood clot pairs experiencing the opposite phenomenon, indicating a decrease in lysis time for longer blood clots. These ratios vary between 0.963 and 0.980.

Overall, TNK reaches greater time ratios than ATP when comparing blood clot pairs from the same scenarios.

5. DISCUSSIONS

The results obtained demonstrate the importance certain properties have during blood clot lysis.

5.1. Therapeutic Drug Injected

Figure 4.4.1 indicates a clear increase in lysis time when injecting TNK instead of ALT. On average, TNK takes 2.5 times longer to dissolve the same blood clot as ATP, with a maximum of 3.6 times more when comparing T08 with A08. The following analysis examines the mechanisms underlying these differences.

The concentration profiles of each drug in the system (see Figure 4.1.2 (a) and (b)) reflect their respective injection regime. ATP's concentration increases sharply, plateaus above $0.025\mu\text{M}$ for about one hour before decreasing exponentially. In contrast, TNK's concentration peaks at $0.085\mu\text{M}$ directly after its injection but decreases drastically soon after. Comparing both profiles, ATP's effect lasts longer than TNK's, but, as TNK's peak rises above ATP's plateau, it will have a stronger impact. ATP reaches a maximum of $0.028\mu\text{M}$ an hour into the process, while TNK crosses this limit half an hour after its injection. Therefore, there is a time window of about 30 minutes where ATP's concentration is greater than TNK's in the system, supporting this hypothesis on the intensity and duration of both drugs.

When comparing the extent-of-lysis graphs for the same lysis scenarios, with only a difference in the drug injected (see Figures 4.1.3 (a) and (b)), the curve gradient for ATP is always smaller than when injecting TNK. Furthermore, its curve often follows a linear to logarithmic profile, depending on the blood clot scenario, while TNK's tends to be more exponential. Therefore, a difference in the lysis process is observable between the two drugs, as ATP's rate increases throughout the simulation, in contrast to TNK, whose effects slow down over time. These mechanisms can be explained by the dosage regime of each drug. As previously explained, ATP has longer-lasting effects, while TNK impacts the blood clot strongly at first but decreases rapidly as well. However, TNK's lysis rate profile becomes

advantageous when dissolving smaller blood clots or during early phases of their development: further analysis on this phenomenon is described in section 5.4.

Figures 4.1.4 (a) and (b) show that the front and rear shells experience the most resistance due to their high fibrin content. ATP starts the rear shell dissolution early in the process, but its lysis rate for each shell is slower than TNK. In contrast, TNK starts the rear shell lysis quite late in the process. This demonstrates two major phenomena differentiating the drugs: ATP reaches and interacts with fibrin at the rear of the blood clot earlier than TNK, implying an ease to penetrate the thrombi. On the other hand, TNK dissolves the fibrin-rich shell more quickly than ATP. Both phenomena can be explained by the drugs' fibrin affinity: TNK, having a higher fibrin affinity, easily gets stuck in the fibrin mesh as soon as it enters the blood clot, when ATP spreads more across the clot (Adivitiya & Khasa, 2017). This high fibrin affinity eases the dissolution of fibrin-rich areas, such as the shells. Many studies on extracted blood clots, such as Parsons et al. (2012), conclude that TNK has a better thrombolysis performance than ATP: this difference is due to real thrombi structures, which usually don't obstruct the entire blood vessel, and therefore have a fibrin-rich shell on their top as well. This increases the fibrin fraction in the blood clot, advantaging TNK in real-life scenarios.

5.2. Blood Clot Type

The observed discrepancies in lysis among thrombus types can be attributed to a combination of their structural properties and the haemodynamic of the vessels in which they form.

Firstly, the significantly longer lysis times predicted for venous clots is more attributable to venous flow conditions than to clot structure. The venous pressure gradient is significantly lower than the arterial pressure drop across clot length, at 10 mmHg/cm and 60 mmHg/cm, respectively (Widrich & Shetty 2024). In fact, despite the theoretical high permeability suggested by the model (refer to Table 3.2.2), the low driving pressure prevents the drug from efficiently accessing binding sites. This is further validated by Piebalgs and Xu (2015)'s study, in which emphasis is put on pressure as a key driving factor for drug transport to target fibrinolytic sites.

Secondly, experimental data suggests that venous clot may require a higher drug dosage to achieve efficient clot lysis due to the vein's larger geometry (Grewal et al., 2022). However, this increase in dosage must be balanced between thrombolytic efficiency and risk of systemic side-effects, as a higher tPA dosage has been shown to increase intracerebral haemorrhage among other risks (Piebalgs et al., 2018).

Additionally, venous thrombi exhibit a higher amount of resistance spread across their entire length, symptomatic to a homogenous structure where fibrin is spread equally all over the clot. ATP's uniform lysis pattern likely reflects the drug's low fibrin affinity, discussed in section 5.1. This causes the drug to be more distributed throughout the clot, leading to a relatively uniform lysis once penetration occurs. In arterial thrombi, Figure 4.2.2 (b) shows that the most resistance

was found to be in the shells of the clot, which is where a higher content of fibrin and lower voidage is exhibited (see Table 3.2.1). This is in accordance with literature, where fibrin is the main component that creates a barrier for the drug to penetrate and dissolve (Di Meglio et al., 2019).

Comparing arterial clots lysis, "red" thrombi are overall lysed quicker than "white" thrombi due to the latter's high outer shell density, exhibiting only 15% of voidage against 50% for "red" thrombi. Despite exhibiting a higher porosity in the core, with 93% voidage, accounting for 90% of the whole clot, the dense platelet-rich shell of the "white" thrombi significantly hindered drug transport to fibrinolytic sites. ATP's extent-of-lysis profiles further explains this mechanism. Indeed, while some of the agent manage to penetrate the permeable core of the clot, the rest of the active drug is trapped behind, still trying to break down the platelet-rich layer of the clot (Mereuta et al. 2021). "Red" thrombi, by contrast, have a more moderate voidage along their entire length. This allows ATP to diffuse more readily through the clot's shell and core.

Additionally, the higher RBC content in "red" thrombi reduces hindrance to the drug transport as the clot breaks down (Weisel & Litvinov, 2021). This is because the RBCs trapped in the fibrin mesh are released during lysis, increasing local permeability and facilitating blood flow through the clot (see Equation 3.1.7). In contrast, "white" thrombi release fewer components as platelets remain embedded within fibrin as the clot degrades. This means that the structural resistance of the shell is maintained for longer, hindering drug penetration.

While "white" clot dissolution was overall longer than "red" thrombi for both agents, TNK shows more consistent lysis pattern across all arterial clots. This is likely due to TNK's high fibrin-binding specificity and longer half-life, as discussed in section 5.1.

5.3. Blood Clot Position

The increased lysis time observed in M2 clots is primarily due to the additional 7.24 mm the drug must travel to reach the clot surface, resulting in a time delay. Because these clots are located farther downstream in the vessel, the drug takes longer to reach its intended target site (Piebalgs et al., 2018).

Additionally, as the agent travels from the injection point to the clot surface, it may react with plasminogen in the plasma, reducing the concentration of active drug needed to dissolve the clot front effectively. For instance, the decrease in plasminogen concentrations exhibited in Figure 4.3.33 Figures 4.3.3 (a) and (b) validates that some reactions occur between the agent and plasminogen. This phenomenon is not observed with TNK because its activation requires direct fibrin contact, which isn't possible before reaching the clot surface (refer to section 5.1) (Szabo et al., 2002). Hence, the agent doesn't undergo premature consumption ahead of the clot, and its lysis time is only delayed by the additional distance to reach the clot's surface. This suggests that TNK has more consistent and predictable performance, regarding thrombi's location, compared with ATP.

Finally, the counterintuitive results in which ATP lysed the core and distal shell faster than the front shell in M2-positioned “white” and venous thrombi can be attributed to the model's limitations regarding drug transport and biochemical reactions. In fact, studies indicate that rear-clot lysis becomes feasible only after recanalization, when most of the clot has already been dissolved creating a pathway for the drug to reach the remaining obstructed region (Piebalgs & Xu, 2015). This pattern is furthermore not observed in “red” thrombi, suggesting that clot permeability and its complex structure further affect downstream thrombolysis.

5.4. Blood Clot Length

Figure 4.4.1 shows that an increase in blood clot length doesn't have a major impact on the lysis duration, but slight variations are still visible. A greater lysis duration is expected for longer clots, as they involve more fibrin that needs to be dissolved by the drug.

This is the case for 5 blood clot pairs, listed in Figure 4.4.2, where the lysis time ratio of biggest over smallest thrombus is greater than 1, reaching 1.124. In all these scenarios, the blood clots are located at the blood vessel inlet. The extent-of-lysis profiles shown in Figures 4.1.3 (a) and (b), discussed earlier, are consistent with these results: as their gradient is positive, lengthening the blood clot leads to a longer lysis time (Seners et al., 2019).

However, as shown in Figure 4.4.3, 5 blood clot pairs experience the opposite phenomenon, where the longer thrombus takes less time to dissolve than the shorter one, although containing more fibrin. This behaviour is explained by tPA having more time to react with the proteins in the plasma before reaching the blood clot. Therefore, the plasma composition varies greatly between the blood vessel inlet and the thrombus, leading to these results. This mechanism was further explained in section 5.3.

TNK is more impacted by the lengthening of the blood clots than ATP, as the time ratios for pairs involving TNK were always greater than ATP's for the same scenarios, as seen on Figures 4.4.2 and 4.4.3. As mentioned in section 5.1, TNK has a better lysis performance than ATP in the blood clot shell but poorer in its core due to its higher fibrin affinity. As the shell-to-total-blood-clot lengths ratio is constant, the lengthening impacts mostly the core of the structure, disadvantaging TNK. However, it dissolves shorter blood clots more efficiently than ATP, when in an early development stage, for instance.

6. CONCLUSION

In this study, a detailed PKPD model was used to investigate how factors such as different therapeutic drug and clot's characteristics related to type, length and position, influenced thrombolytic efficiency.

It was concluded that TNK's lysis pattern tended to be more consistent and more effective at the thrombus outer shell due to high fibrin specificity while ATP dissolved the overall blood clots faster because it penetrates the core more easily. Furthermore, the simulations showed that “red” arterial clots were the easiest to get lysed because of their higher permeability, whereas venous clots were more resistant due to unfavourable haemodynamic in veins. It was noted that

unexpected behaviours were displayed for longer and distal clots which experienced lower lysis time when the clot was positioned further away from the inlet. Overall, these findings suggest that thrombolytic success depends on the complex interplay between clot properties and drug's transport and kinetic behaviour.

Although insightful, some model assumptions could further be relaxed in future work to make predictions more clinically relevant. These include developing a more realistic blood vessel geometry model or introducing a time-varying pressure gradient. Additionally, collecting data on venous clots would improve the results accuracy and bring new insights into blood clot lysis.

7. ACKNOWLEDGEMENTS

Special thanks are extended to Gots Narongchartsopon, the GTA for this project, for offering continuous guidance throughout the study.

8. REFERENCES

- Adivitiya and Khasa, Y.P. (2017) The evolution of recombinant thrombolytics: Current status and future directions. *Bioengineered*, 8(4), 331-358. doi:10.1080/21655979.2016.1229718
- Banerjee, A., Chisti, Y. and Banerjee, U.C. (2004). Streptokinase—a clinically useful thrombolytic agent. *Biotechnology Advances*, [online] 22(4), pp.287–307. doi:10.1016/j.biotechadv.2003.09.004
- Bannish, B.E. and Hudson, N.E. (2021). The utility and potential of mathematical models in predicting fibrinolytic outcomes. *Current Opinion in Biomedical Engineering*, 20, p.100337. doi:10.1016/j.cobme.2021.100337
- Chen, L., Zhan, Q., Ma, C., Liu, Q., Zhang, X., Tian, X., Jiang, Y., Dong, Y., Chen, S. and Lu, J. (2015) Reproducibility of Middle Cerebral Artery Stenosis Measurements by DSA: Comparison of the NASCET and WASID Methods. *PLoS ONE*, 10(6), e0130991. doi:10.1371/journal.pone.0130991
- Chernysh, I.N., Nagaswami, C., Kosolapova, S., Peshkova, A.D., Cuker, A., Cines, D.B., Cambor, C.L., Litvinoc, R.I. and Weisel, J.W. (2020) The distinctive structure and composition of arterial and venous thrombi and pulmonary emboli. *Scientific Reports*, 10, 5112. doi:10.1038/s41598-020-59526-x
- Di Meglio, L., Desilles, J.-P., Ollivier, V., Nomenjanahary, M.S., Di Meglio, S., Deschildre, C., Loyau, S., Olivot, J.-M., Blanc, R., Piotin, M., Bouton, M.-C., Michel, J.-B., Jandrot-Perrus, M., Ho-Tin-Noé, B. and Mazighi, M. (2019) Acute ischemic stroke thrombi have an outer shell that impairs fibrinolysis. *Neurology*, [online] 93(18), pp.e1686–e1698. doi: 10.1212/wnl.00000000000008395
- Diamond, S.L. (1999) Engineering Design of Optimal Strategies for Blood Clot Dissolution. *Annual Review of Biomedical Engineering*, 1(1), pp.427–461. doi:10.1146/annurev.bioeng.1.1.427
- Grewal, S.K., Hedrick, A.L., Man, L., Sharma, A.M., Desai, K.R. and Khaja, M.S. (2022) A Brief Review of Thrombolytics for Venous Interventions. *Seminars in Interventional Radiology*, 39(04), pp.394–399. doi:10.1055/s-0042-1757318

- Gu, B., Piebalgs, A., Huang, Y., Roi, D., Lobotesis, K., Longstaff, C., Hughes, A.D., Chen, R., Thom, S.A. and Xu, X.Y (2019a) Computational simulations of thrombolysis in acute stroke: Effect of clot size and location on recanalisation. *Medical Engineering and Physics*, 73, pp.9-17. doi:[10.1016/j.medengphy.2019.07.014](https://doi.org/10.1016/j.medengphy.2019.07.014)
- Gu, B., Piebalgs, A., Huang, Y., Longstaff, C., Hughes, A.D., Chen, R., Thom, S.A. and Xu, X.Y (2019b) Mathematical Modelling of Intravenous Thrombolysis in Acute Ischaemic stroke: Effects of Dose Regimens on Levels of Fibrinolytic Proteins and Clot Lysis Time. *Pharmaceutics*, 11, 111. doi:[10.3390/pharmaceutics11030111](https://doi.org/10.3390/pharmaceutics11030111)
- Jolugbo, P. and Ariëns, R.A.S. (2021). Thrombus Composition and Efficacy of Thrombolysis and Thrombectomy in Acute Ischemic Stroke. *Stroke*, 52(3). doi:[10.1161/strokeaha.120.032810](https://doi.org/10.1161/strokeaha.120.032810)
- Menon, B.K., Buck, B.H., Singh, N., Deschaintre, Y., Almekhlafi, M.A., Coutts, S.B., Thirunavukkarasu, S., Khosravani, H., Appireddy, R., Moreau, F., Gubitz, G., Tkach, A., Catanese, L., Dowlatsahi, D., Medvedev, G., Mandzia, J., Pikula, A., Shankar, J., Williams, H. and Field, T.S. (2022) Intravenous tenecteplase compared with alteplase for acute ischaemic stroke in Canada (AcT): a pragmatic, multicentre, open-label, registry-linked, randomised, controlled, non-inferiority trial. *The Lancet*, [online] 400(10347), pp.161–169. doi:[10.1016/S0140-6736\(22\)01054-6](https://doi.org/10.1016/S0140-6736(22)01054-6)
- Mereuta, O.M., Agarwal, T., Ghozy, S., Dai, D., Arul, S., Brinjkiji, W., Kallmes, D.F. and Kadirvel, R. (2024) Shell versus core architecture and biology of thrombi in acute ischemic stroke: A systematic review. *Journal of Stroke and Cerebrovascular Diseases*, 33(1), 106–115. doi:[10.1177/10760296231213632](https://doi.org/10.1177/10760296231213632)
- Mereuta, O.M., Rossi, R., Douglas, A., Gil, S.M., Fitzgerald, S., Pandit, A., McCarthy, R. and Gilvarry, M. (2021) Characterization of the ‘White’ Appearing Clots that Cause Acute Ischemic Stroke. *Journal of Stroke and Cerebrovascular Diseases*, 30, 12. doi:[10.1016/j.jstrokecerebrovasdis.2021.106127](https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.106127)
- Narongchartsopon, G.S., Tan, Y.J., Yang, Y. and Xu, X.Y (2025) A computational study of clot composition and heterogeneity on thrombolysis. *SSRN*
- Narwal, A., Whyte, C.S. and Mutch, N. (2023) Location, location, location: Fibrin, cells, and fibrinolytic factors in thrombi. *Front. Cardiovasc. Med*, 9, 1070502. doi:[10.3389/fcvm.2022.1070502](https://doi.org/10.3389/fcvm.2022.1070502)
- Nikitin, D., Choi, S., Mican, J., Toul, M., Ryu, W.-S., Damborsky, J., Mikulik, R. and Kim, D.-E. (2021) Development and Testing of Thrombolytics in Stroke. *Journal of Stroke*, 23(1), pp.12–36. doi:[10.5853/jos.2020.03349](https://doi.org/10.5853/jos.2020.03349)
- Peters, N.A. and Paciullo, C.A. (2015) Alteplase for the Treatment of Pulmonary Embolism. *Advanced Emergency Nursing Journal*, 37(4), pp.258–272. doi:[10.1097/tme.0000000000000082](https://doi.org/10.1097/tme.0000000000000082)
- Piebalgs, A., Gu, B., Roi, D., Lobotesis, K., Thom, S. and Xu, X.Y (2018) Computational Simulations of Thrombolytic Therapy in Acute Ischaemic Stroke. *Scientific Reports*, 8(1), p.15810. doi:[10.1038/s41598-018-34082-7](https://doi.org/10.1038/s41598-018-34082-7)
- Piebalgs, A. and Xu, X.Y. (2015) Towards a multi-physics modelling framework for thrombolysis under the influence of blood flow. *Journal of The Royal Society Interface*, 12(113), pp.20150949–20150949. doi:[10.1098/rsif.2015.0949](https://doi.org/10.1098/rsif.2015.0949)
- Rohan, V., Baxa, J., Tupy, R., Cerna, L., Sevcik, P., Friesl, M., Polivka, J., Polivka, J. and Ferda, J. (2014) Length of Occlusion Predicts Recanalization and Outcome After Intravenous Thrombolysis in Middle Cerebral Artery Stroke. *Stroke*, 45(7), pp.2010-2017, doi:[10.1161/STROKEAHA.114.005731](https://doi.org/10.1161/STROKEAHA.114.005731)
- Rossi, R.; Fitzgerald, S.; Gil, S. M.; Mereuta, O. M.; Douglas, A. and Pandit, A. (2021) Correlation between acute ischaemic stroke clot length before mechanical thrombectomy and extracted clot area: Impact of thrombus size on number of passes for clot removal and final recanalization. *European Stroke Journal*, 6(3), pp.254-261. doi:[10.1177/23969873211024777](https://doi.org/10.1177/23969873211024777)
- Sidik, W.A. and Mazumdar, J.N. (1994) A mathematical study of turbulent blood flow through an arterial bifurcation. *Australas Phys. Eng. Sci. Med.*, 17(1), pp.1-13. PMID: 8198503
- Seners, P., Delepierre, J., Turc, G., Henon, H., Piotin, M., Arquizan, C., Cho, T.-H., Lapergue, B., Cottier, J.-P., Richard, S., Legrand, L., Bricout, N., Mazighi, M., Dargazanli, C., Nighoghossian, N., Consoli, A., Debiais, S., Bracard, S., Naggara, O., Leclerc, X., Obadia, M., Costalat, V., Berthezène, Y., Tisserand, M., Narata, A.-P., Gory, B., Mas, J.-L., Oppenheim, C., Baron, J.-C. (2019) Thrombus Length Predicts Lack of Post-Thrombolysis Early Recanalization in Minor Stroke With Large Vessel Occlusion. *Stroke*, 50(3), 761-764. doi:[10.1161/STROKEAHA.118.023455](https://doi.org/10.1161/STROKEAHA.118.023455)
- Szabo, S., Letsch, R., Ehlers, R., Walter, T., Kazmaier, S., Helber, U. and Hoffmeister, H.M. (2002) Absence of paradoxical thrombin activation by fibrin-specific thrombolytics in acute myocardial infarction. *Thrombosis Research*, 106(2), pp.113–119. doi:[10.1016/s0049-3848\(02\)00084-1](https://doi.org/10.1016/s0049-3848(02)00084-1)
- touchNEUROLOGY (2025) Tenecteplase receives FDA approval for acute ischemic stroke - touchNEUROLOGY. [online] Available at: <https://touchneurology.com/insight/tenecteplase-receives-fda-approval-for-acute-ischemic-stroke/> [Accessed 10 Dec. 2025].
- Weisel, J.W. and Litvinov, R.I. (2021) Visualizing thrombosis to improve thrombus resolution. *Research and Practice in Thrombosis and Haemostasis*, 5(1), pp.38–50. doi:[10.1002/rth2.12469](https://doi.org/10.1002/rth2.12469)
- Widrich, J. and Shetty, M. (2024) ‘Physiology, Pulmonary Vascular Resistance’. In: *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK554380/> (Accessed: 9 December 2025).
- World Health Organization (2025) Cardiovascular diseases (CVDs). [online] *World Health Organization*. Available at: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).
- Yang, Y., Gu, B. and Xu, X.Y (2023) In Silico Study of Different Thrombolytic Agents for Fibrinolysis in Acute Ischemic Stroke. *Pharmaceutics*, [online] 15(3), pp.797–797. doi:[10.3390/pharmaceutics15030797](https://doi.org/10.3390/pharmaceutics15030797)

Bayesian Optimization in Latent Variable Models for Expensive Black-Box Problems

Kathiir Sivanehru and Simran Gadhia

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Many problems within chemical engineering are modelled as black-box functions, meaning derivative-based techniques fail because gradients are unavailable, expensive to approximate, or unreliable. This motivates surrogate-based, sample-efficient methods such as Bayesian optimisation (BO), although its performance deteriorates in high dimensions. To address this, the paper combines BO with latent variable models (LVMs): principal component analysis (PCA), autoencoders and variational autoencoders (AEs and VAEs). These compress high-dimensional problems into a lower latent space where optimisation is more feasible. Sequential domain reduction (SDR) trust-region mechanisms are also incorporated to prevent wandering and improve stability. The effectiveness of this pipeline depends on the latent model's ability to encode and decode reliably so that structure in the original function is preserved. This paper adopts a tiered framework of increasing realism: tier 1 uses synthetic 20D functions to establish a benchmark; tier 2 introduces structured nonlinear manifolds; and tier 3 considers a simple reactor case with a multidimensional input. After training LVMs and tuning BO across tiers 1 and 2, VAEs emerged as the most robust across objectives. In the reactor study, the final model learnt the intrinsic manifold with reconstruction errors of order 10^{-6} , enabling the full pipeline (BO+VAE+SDR) to achieve a tenfold improvement in maximum yield. Overall, the results show that BO in a latent space performs far better than in its native dimensions, with SDR further accelerating convergence.

1 Introduction

As industrial and agricultural demands continue to rise, the need for chemical processes that turn raw materials into useful products becomes more important. When this is combined with tighter budgets and pressure to reduce environmental impact, optimisation naturally moves to the forefront.

In chemical engineering, reactor optimisation is usually carried out using CFD, where large sets of PDEs are solved to model flow, mixing and reaction environments in complex 3D domains [1, 2]. These simulations are essential, but they behave as black-box functions: gradients are unavailable or unreliable, the internal steps are non-smooth, and standard derivative-based solvers struggle as a result [3, 4].

This has pushed engineers towards derivative-free optimisation. However, many classical approaches become impractical once the dimension grows, and evaluating thousands of points is rarely realistic for CFD-level costs. Even surrogate-based methods such as Bayesian optimisation (BO) begin to deteriorate in high dimensions: acquisition functions flatten, kernel length scales lose meaning, and the Gaussian process becomes harder to train [5, 6, 7]. These issues are important because many realistic design problems easily exceed 30–40 variables [8].

One way around this is to reduce the dimensionality

first. PCA provides a linear reduction but cannot represent the curved or nonlinear structure typically seen in chemical engineering data [9]. Autoencoders (AEs) extend this by learning nonlinear manifolds through an encoder–decoder mapping [10]. Variational autoencoders (VAEs) add further structure by enforcing a smoother latent geometry, which generally avoids sudden jumps or irregular regions during optimisation [11]. These models only help BO if the latent space preserves the important features of the objective; if it folds or distorts key directions, the optimiser may be misled [12].

To study when latent-space BO remains reliable, this paper uses a tiered sequence of problems. Tier 1 examines BO on common 20D benchmark functions (Ackley, Rosenbrock, Levy, Rastrigin) to understand behaviour on multimodal or ridge-dominated surfaces. Tier 2 introduces nonlinear embeddings to construct structured manifolds so the latent models have something meaningful to learn, following the ideas of Wang et al. [5]. Tier 3 applies the full pipeline to a simple reactor model, where the physical inputs appear low-dimensional but expand to 36 variables once discretised for simulation.

The aim is to determine when latent-space BO offers a genuine improvement over native BO in high-dimensional settings, and how region-refinement strategies such as sequential domain reduction (SDR)

can reduce exploratory drift. These ideas are tested across the three tiers outlined above. The remainder of the paper presents the background, methodology, results and conclusions.

2 Background

2.1 Gaussian Process

Before explaining BO, it is important to highlight the key role that Gaussian processes (GPs) play. They are the standard non-parametric model for black-box optimisation problems because they provide both a prediction and an uncertainty at any input. They make no assumption about a fixed parametric form, which makes them well suited to model functions flexibly and reliably even when only limited data are available. GPs also offer calibrated uncertainty and closed-form posterior updates, giving reliable performance in demanding engineering simulations [13, 14]. A practical limitation is the $O(N^3)$ training cost [13], which strengthens the case for sample-efficient methods such as BO.

The initial dataset is

$$D_0 = (z_i, y_i)_{i=1}^N, \quad (1)$$

and before any data are observed, the GP assumes a prior

$$f(z) \sim \mathcal{GP}(0, k(z, z')), \quad (2)$$

with the kernel encoding assumptions about smoothness and structure. Conditioning this prior on the dataset gives the posterior

$$f | D_t \sim \mathcal{GP}(\mu_t(z), k_t(z, z')). \quad (3)$$

Kernel choice is problem-dependent. The RBF kernel assumes very smooth behaviour. In contrast, the Matérn-5/2 kernel is better suited to rougher, multi-modal objectives and is therefore widely used in BO:

$$k_{\nu=2.5}(r) = \sigma_f^2 \left(1 + \sqrt{5} r + \frac{5}{3} r^2 \right) \exp(-\sqrt{5} r), \quad (4)$$

$$r^2 = \sum_{i=1}^d \frac{(z_i - z'_i)^2}{\ell_i^2}.$$

ARD extends this by giving each latent coordinate its own lengthscale so the GP can reduce the influence of irrelevant or noisy directions.

After each function evaluation, the GP updates to give the posterior mean,

$$\mu_t(z) = k_{\tau}(K + \sigma_n^2 I)^{-1} y, \quad (5)$$

and posterior variance,

$$\sigma_t^2(z) = k(z, z) - k_{\tau}(K + \sigma_n^2 I)^{-1} k_{\tau}, \quad (6)$$

which together guide the BO acquisition step. In this work, all GP modelling is carried out in latent space, so the behaviour of the kernel depends directly on the geometry learned by the PCA, AE or VAE encoders.

2.2 Bayesian Optimisation + Sequential Domain Reduction

BO aims to solve the following optimisation problem:

$$x^* = \arg \min_{x \in \mathcal{X}} f(x), \quad (7)$$

and is especially useful when full evaluations of f are expensive and derivative free [15]. In practice, BO does not assume access to the true function directly, so each evaluation is modelled with noise:

$$y = f(x) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_n^2), \quad (8)$$

BO consists of three main components: an initial dataset, a GP surrogate (Section 2.1), and an acquisition function [16]. These interact in a simple loop: The GP is first trained on the current dataset; the acquisition function then scores possible inputs based on how promising they look; the best candidate is evaluated and added to the dataset.

Acquisition functions measure the potential utility of sampling a point by balancing exploitation and exploration [16]. A common example is the Upper Confidence Bound (UCB),

$$\alpha_{\text{UCB}}(z) = \mu(z) + \kappa \sigma(z), \quad (9)$$

where κ controls the exploration–exploitation balance. [17]. Another widely used choice is Expected Improvement (EI), which adapts naturally to uncertainty and has shown strong empirical performance [15]. It is given by:

$$\text{EI}(z) = (f_{\text{best}} - \mu(z)) \Phi\left(\frac{f_{\text{best}} - \mu(z)}{\sigma(z)}\right) + \sigma(z) \phi\left(\frac{f_{\text{best}} - \mu(z)}{\sigma(z)}\right), \quad (10)$$

Here, the first term represents exploitation and the second exploration.

Once a new point is chosen, $f(x^*)$ is evaluated and added to the dataset, and the GP is updated, until budget is expended. In this work, BO is applied either in the original space (native) x or in the latent space z created by PCA, AEs or VAEs, so the acquisition function is optimised on z and the chosen point is then decoded before the true objective is evaluated.

BO is still prone to exploring unproductive regions since no encoder/decoder model is perfect, so another angle of dimension reduction is used alongside LVMs, which is SDR. This constrains BO to a local region which adapts based off of BOs current progress, similar to trust region approaches seen in Eriksson et al. SDR adapts a trust region around the current best point, expanding after improvement and contracting otherwise[6].

2.3 Latent Variable Models

While BO provides a sample-efficient approach, its performance deteriorates sharply as dimensionality increases, motivating the use of dimensionality reduction. PCA is the most classical approach for this, identifying the direction of maximum variance, but it is restricted to linear surfaces. This is due to distorting nonlinear manifolds in a way that misidentifies directions of variance. Most chemical engineering problems have complicated landscapes (multimodal, ridge-dominated) which are not translated well[9]. To more accurately capture non-linear features, AEs are introduced, which learn a non-linear low dimensional representation using neural networks [10]. An autoencoder consists of a two parts: an encoder that maps an input to a latent variable,

$$z = f_{\phi}(x), \quad (11)$$

and a decoder that reconstructs back to the input,

$$\hat{x} = g_{\theta}(z). \quad (12)$$

The network is trained by minimising a mean squared reconstruction error (MSE loss),

$$\mathcal{L}_{AE} = \|x - \hat{x}\|^2, \quad (13)$$

so the latent space is forced to capture the structure needed to recover the original data, so they can cover nonlinearities.

However, AEs produce a latent space discretely via a vector, which is not fully smooth for BO to explore. VAEs address this by mapping to a latent distribution. This Gaussian distribution produces a smoother, more predictable space that better aligns with the assumptions required for BO, establishing the theory behind VAEs being the most suitable LVM for BO.

Instead of defining single latent vectors z the VAE outputs both a mean vector and log variance vector, which defines a multivariate Gaussian in the latent space:

$$q_{\phi}(z | x) = \mathcal{N}\left(z; \mu(x), \text{diag}(\sigma^2(x))\right) \quad (14)$$

Since each input is represented by a cloud of possible latent points rather than a fixed point, the space is more continuous and better behaved. To sample from the above distribution, the reparametrisation trick is used to avoid the use of a purely random sample which would break the mathematical chain needed for training [11] [18]:

$$z = \mu(x) + \sigma(x) \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \quad (15)$$

After sampling, the decoder provides a reconstruction the same way an AE does. The new loss term becomes:

$$\mathcal{L}_{VAE} = \mathbb{E}_{q_{\phi}(z|x)} [\log p_{\theta}(x | z)] - \text{KL}(q_{\phi}(z | x) \| p(z)), \quad (16)$$

with both the reconstruction term measuring how well the decoder produces x from the sampled z , and the Kullback-Leibler (KL) divergence term which pushes the distribution back towards a simple normal. This ensures smoothness and good geometry, resulting in BO being able to make small moves in the latent space, giving small predictable changes upon decoding, making the GP surrogate much more reliable. It is important to note this 'KL annealing' is delayed so the model can first learn structure before being adjusted. Figure 1 shows the neural network present in a VAE:

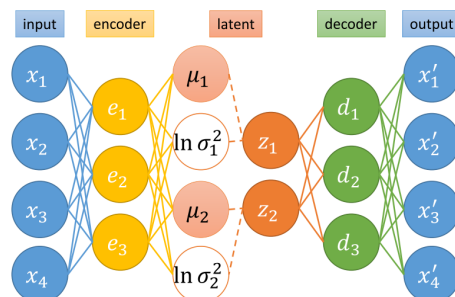


Figure 1: Schematic of a variational autoencoder, showing encoder outputs μ and $\log \sigma^2$, sampling to obtain z , and decoder reconstruction $\hat{x}$. Reproduced from Portillo et al. [19].

3 Methodology

The workflow for this paper can be split into 3 sections, each with increasing realism to practical chemical engineering problems:

- Tier 1: Synthetic 20D functions with no intrinsic structure. These are used to baseline LVM performances and discover optimal hyper parameters for BO/AE.
- Tier 2: Synthetic 20D functions with an underlying structure to better represent real problems.

- Tier 3: Application to a simulated chemical reactor with the optimisation objective of maximising yield.

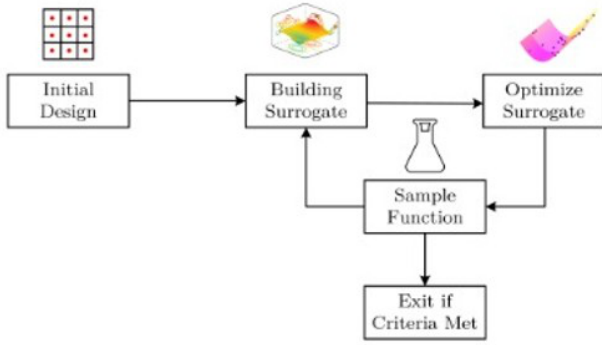


Figure 2: Fundamental workflow of model-based DFO, taken from Mathias et al. [20].

Algorithm 1 Latent space Bayesian optimisation with SDR

```

1: Given
   Design space  $\mathcal{X}$  and objective  $f$ 
   Latent model  $\mathcal{M}$  (PCA, AE, or VAE)
   Evaluation budget  $B$ 
2: Initialisation
3: Sample initial  $\{x_i\}_{i=1}^{N_0}$  and evaluate  $y_i = f(x_i)$ 
4:  $D_0 \leftarrow \{(x_i, y_i)\}_{i=1}^{N_0}$ 
5: Train  $\mathcal{M}$  and freeze  $f_\phi, g_\theta$ 
6: Encode  $z_i \leftarrow f_\phi(x_i)$  and fit GP on  $(z_i, y_i)$ 
7:  $t \leftarrow 0, B_{\text{used}} \leftarrow N_0$ 
8: Initialise SDR region  $\mathcal{Z}_0$ 
9: while  $B_{\text{used}} < B$  do
10:   Acquisition
11:   Generate candidates  $\{z_j^{(c)}\}_{j=1}^{N_c} \subset \mathcal{Z}_t$  using Sobol sampling
12:   for  $j = 1$  to  $N_c$  do
13:     Compute  $\text{EI}(z_j^{(c)})$ 
14:   end for
15:    $z^* \leftarrow \arg \max_j \text{EI}(z_j^{(c)})$ 
16:   Decode and evaluate
17:    $x^* \leftarrow g_\theta(z^*)$ 
18:    $y^* \leftarrow f(x^*)$ 
19:   Update
20:    $D_{t+1} \leftarrow D_t \cup \{(z^*, y^*)\}$ 
21:   Refit GP on  $D_{t+1}$ 
22:    $B_{\text{used}} \leftarrow B_{\text{used}} + 1$ 
23:    $t \leftarrow t + 1$ 
24:   Update SDR region  $\mathcal{Z}_t$ 
25: end while
26: return best  $x$  observed in  $\{(x_i, y_i)\}$ 

```

Figure 2 and Algorithm 1 above show the general optimisation loop used in the paper and the full pipeline

loop, respectively.

- Initial design: Samples to generate initial picture of objective function. Tier 1 and 2 sample a hypercube in their respective dimensions, tier 3 constructs a noisy grid of combinations of the inputs.
- Building surrogate: Use data from initial design to build a surrogate model. All tiers use a GP surrogate built in similar manners.
- Optimise surrogate: Find next best points to sample. The acquisition function employed by all tiers is EI.
- Sample function: Evaluate the true function. In tier 1/2, this is the synthetic function, and in tier 3 this is the ODE solver of the reactor simulation.
- Exit if criteria is met: Termination, all tiers are capped to 100 iterations.

3.1 Tier 1

3.1.1 Test functions, Design Space and Preprocessing

The test functions used are Rosenbrock, Ackley, Levy and Rastrigin. All are defined over a known global minimum, allowing consistent benchmarking. The input dimension was fixed at 20 to mimic the high dimensionalities seen in real chemical engineering problems, with the search domain taken as the standard hypercube $[-5, 5]^{20}$. A total of 10000 points were generated to create an intentionally unstructured dataset, providing a baseline for LVM performance when no underlying manifold exists. Initial points were sampled using a Sobol sequence before evaluating each test function.

Once the data was generated, each coordinate was standardised to zero mean and unit variance to maintain consistent scaling for all subsequent decoding and BO steps.

3.1.2 LVM Training

Each LVM is trained on a 80/20 training/validation split, with early stopping taking effect when there are minimal improvements to validation reconstruction error for a given patience of 100.

The first LVM which is fit to this processed dataset is the PCA. This was tested on all latent dimension inclusive between 2-8. PCA is used to provide a linear baseline for BO in reduced space.

Then, AEs are trained. The same dataset is used, with the architecture consisting of a fully connected encoder and decoder.

A variety of activation functions were tested: ReLu, SiLu, and tanh. SiLu was chosen since it performs

the best over many epochs and it gave the most stable results. However, to avoid hand-tuning the other parameters of the AE, Optuna was used [21]. Each Optuna trial constructing a different AE with different permutations of parameters and applying each AE to a 80/20 data split, which was evaluated with MSE. The final hyperparameters are included in Table 1

Table 1: Key hyperparameters for AE training.

Hyperparameter	Value
Architecture	Symmetric encoder/decoder
Layer widths	$D \rightarrow 128 \rightarrow 64 \rightarrow 32 \rightarrow d \rightarrow 32 \rightarrow 64 \rightarrow 128 \rightarrow D$
Activation	SiLU
Batch size	128
Optimiser	Adam
Learning rate	2.07×10^{-3}

After appropriate losses were reached, the encoder and decoder were frozen to be used as latent space translators, preparing them for usage with BO.

The final LVM investigated was the VAE. The same architecture from Table 1 was used to build on the earlier results and avoid repeating the Optuna search. The additional loss term in Equation 16 forces the latent distribution towards a Gaussian, so Optuna was instead applied to the VAE-specific parameters: the initial β , warm-up period, rate of increase and maximum value. This produced an effective KL-annealing schedule that balanced fitting the encoder output to a Gaussian while avoiding posterior collapse, allowing the VAE to first learn the underlying structure before the KL term was fully imposed. The final schedule was: a linear annealing scheme in which $\beta = 0$ for the first 100 epochs, after which it increased by 0.01 per epoch until reaching a maximum value of $\beta = 1$. For reproducibility, the mean of the encoder distribution was used without sampling, since BO requires a deterministic mapping from input space to latent space.

3.1.3 BO + GP

The initial design points are standardised, encoded and whitened to produce the coordinates for GP as per Equation 1. Matern-5/2 with ARD was the kernel of choice, since it consistently outperformed RBF. Kernel parameters are re-fit with each BO iteration via the maximisation of the log marginal likelihood, with a small jitter introduced to ensure numerical stability.

Once the GP has been initialised, Sobol candidates are generated with every iteration. The EI algorithm detailed in Equation 10 automatically balances exploration and exploitation to select the best latent candidate. The frozen decoder then decodes z (inverse-

whiten, decode, de-standardise) so that the true function can be evaluated. This new result is appended to the GP dataset, GP hyperparameters are updated, and iteration continues until budget is exhausted.

3.2 Tier 2

This tier adopts the methodology from Wang et al. [5] to generate synthetic data to better model a high dimensional chemical engineering problems, which shows that high dimensional BO can remain tractable when the objective depends only on a low dimensional linear subspace of the ambient domain. Our structure meets this criteria; instead of treating each dimension as independent, we impose a low rank structure within a high dimensional ambient space via nonlinear embeddings and noise. This generates data that lie on or near lower dimensional manifolds, therefore creating a meaningful latent representation that can be exploited by dimensionality reduction methods, ultimately making tier 2 more realistic.

Let d_Z denote the intrinsic dimension and D the ambient dimension, with $D \gg d_Z$. For each base test function $f : \mathbb{R}^{d_Z} \rightarrow \mathbb{R}$, we generate a high dimensional input $x \in \mathbb{R}^D$ by first sampling a latent vector $z \in \mathbb{R}^{d_Z}$ and then applying a sequence of nonlinear embeddings together with noise. The latent samples z are drawn from a scrambled Sobol design over a bounded box $[\ell_z, u_z]^{d_Z}$.

Given the latent matrix Z , we construct the observed high dimensional data $X \in \mathbb{R}^{N \times D}$ as a concatenation of five groups, where each group represents a different transformation of the same latent vector z :

- **Group 1: Non-linearly warped active latent variables.** First, a smooth monotone nonlinearity that preserves invertibility is applied. Since this block also contains a realistic distortion, it contains the entire optimisation signal.

$$x_{G1}(z) = \alpha \sinh(z), \quad \alpha = 0.5 \quad (17)$$

During evaluation, the latent variable is recovered via:

$$z = \operatorname{asinh}\left(\frac{1}{\alpha} x_{G1}\right) \quad (18)$$

- **Group 2: Linear redundant dimensions.** Engineering systems tend to have linear correlations however, in this study we chose to omit this to focus on the effect of non-linearity in problems.
- **Group 3: Highly nonlinear redundant dimensions.** A richer nonlinear map is applied to better mimic chemical engineering problems by producing high-curvature, multimodal behaviour,

unrelated to the true optimisation landscape.

$$H(z) = A_2 z, \quad A_2 \in \mathbb{R}^{d_Z \times d_{\text{nonlin}}} \quad (19)$$

where the entries of A_2 are drawn from a Gaussian distribution and scaled by $1/d_Z$. The nonlinear redundant block is then defined as

$$x_{G3}(z) = 5 \left[\tanh(H) + 0.2 \sin(2H) + 0.1 H \odot H \right] \quad (20)$$

where $\odot$ denotes element wise multiplication.

- **Group 4: Heteroscedastic correlated noise dimensions.** To introduce realistic heteroscedastic measurement noise, we add noise whose variance depends on the magnitude of the latent variable. Let

$$r(z) = \frac{\|z\|^2}{d_Z}, \quad \sigma_{\text{corr}}(z) = 0.5 + 0.5 r(z) \quad (21)$$

and draw a correlated base term

$$\varepsilon_{\text{corr}}(z) = A_3 z + \eta, \quad \eta \sim \mathcal{N}(0, 0.3^2 I) \quad (22)$$

with A_3 Gaussian and scaled by $1/d_Z$. The correlated noise block becomes

$$x_{G4}(z) = \sigma_{\text{corr}}(z) \varepsilon_{\text{corr}}(z) \quad (23)$$

- **Group 5: Pure noise dimensions.** Finally, for the remaining dimensions, we add high variance independent noise that carries no objective information.

$$x_{G5} = \sigma_{\text{pure}} \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I), \quad \sigma_{\text{pure}} = 2.0. \quad (24)$$

The final ambient input is assembled as

$$X = [G1, G2, G3, G4, G5] + \text{jitter},$$

where a small Gaussian perturbation is added element-wise to mimix measurement jitter. The dimensionality of each group is provided in Table 2 as we apply this embedding to each of the four test functions, with different latent dimensions.

Table 2: Dimensionality of each group used in the synthetic embedding for different latent dimensions.

Group	5D	3D	7D	9D
Nonlinearly warped active latent variables (G1)	5	3	7	9
Linear redundant (G2)	0	0	0	0
Highly nonlinear redundant (G3)	12	14	10	8
Heteroscedastic correlated noise (G4)	1	1	1	1
Pure noise (G5)	2	2	2	2

Therefore with this construction, the true objective only depends on the first group $G1$, and the remaining coordinates come from redundant embeddings and noise. This ensures that although the observable data are high dimensional, the effective dimensionality of the optimisation problem remains equal to d_z .

Repeating what we did in tier 1, we test with perform BO, after reducing dimensions of this new data set with PCA and VAEs for each of the four test functions.

This evaluation enables to understand how these methods behave when the dimensionality reduction is meaningful and the latent manifold faithfully captures the intrinsic structure of the problem.

3.3 Tier 3

Ultimately, tier 3 is the cumulation of all the work above, with theory deployed into real chemical engineering practice. Learnings from tier 1 and 2 aim to translate into the effective maximisation of yield from a simple reactor case study.

3.3.1 Reactor model and data generation

The reactor model is detailed in Figure 3. It consists of two simple kinetic pathways, both governed by first order Arrhenius kinetics:



where both receive true model input variables: temperature, tau and ruta. Although this may seem like a simple 3D problem, the reactor ODE solver requires each input in an expanded form:

1. Temperature - 16D pipe/segment representation
2. Tau τ (residence time) - 16D residence-structure representation
3. Ruta - 4D categorical embedding (which reaction pathway to pursue)

Therefore, the real optimisation domain is 36D, making pure BO ill-posed, hence the need for LVMs. 800 samples are used, created from a noisy grid of physical experiments from 20 temperatures, 20 τ and 2 rutas, with a small Gaussian jitter imposed on τ and ruta. The code for the reactor is adapted from Zavalab et al. [22].

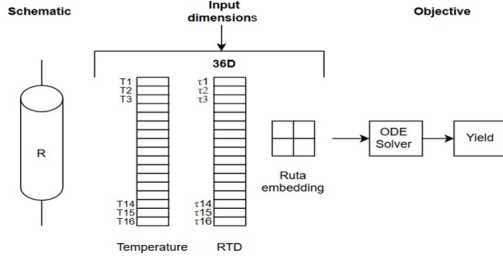


Figure 3: Reactor schematic

3.3.2 Reactor VAE + BO

The only LVM used for the final case was a VAE, since tier 1 and 2 benchmark VAE performance to be the strongest in terms of minimisation and consistency. A latent dimension of 3 is naturally chosen since the true underlying dimension is 3. The encoder architecture was $D \rightarrow 64 \rightarrow 32 \rightarrow d(3) \rightarrow 32 \rightarrow 64 \rightarrow D$ with a SiLu activation, with the same decoder reconstruction. Similar parameters determined from optuna testing from tier 1 and 2 were integrated into tier 3, as well as: latent encoding; bounds; initial BO design; GP surrogate; latent BO loop. SDR was imposed in separate runs to identify its true effect. The objective is the maximisation of yield from the reactor, but to keep graphs and results consistent, a minimisation objective was used (minimise negative yield).

4 Results

4.1 Tier 1

Tier 1 acts as a benchmark for later tiers, with a broad range of findings.

4.1.1 AE vs VAE

Figures 4 and 5 show the difference between AE and VAEs for the four functions. The VAE is found to converge to a lower optima than the AE without exception, despite VAEs suffering slightly worse reconstruction losses consistently (explained later in Figure 11). This proves that BO values a smooth latent space, which is generated in the form of the VAEs continuous encoder output versus the AEs discrete latent vector.

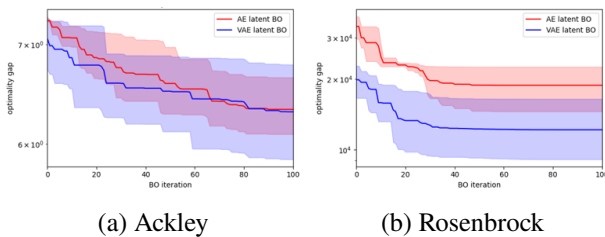


Figure 4: AE and VAE performance on Ackley and Rosenbrock (Tier 1).

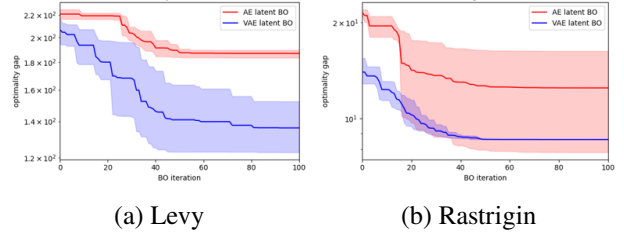


Figure 5: AE and VAE performance on Levy and Rastrigin (Tier 1).

4.1.2 Latent Dimensions

Figures 4 and 5 reinforces the need for a smooth BO space. Figure 6 shows that a larger latent space leads to lower MSE losses, due to the encoder losing less data in compression. One may think that a higher latent dimension leads to a better BO objective, but Figure 7 contradicts this, since as discussed in Section 4.1.1, the BO values a small latent space first. Even if the decoder output is not perfect, the smaller dimension latent spaces perform better as the VAE compresses aggressively, removing redundant or noisy degrees of freedom in the original 20D landscape. This produces a more regular manifold and the BO sees less clutter in the latent space, making it easier to navigate. Specifically, it is easier for the GP kernel to model correlations and the acquisition function to identify promising regions reliably.

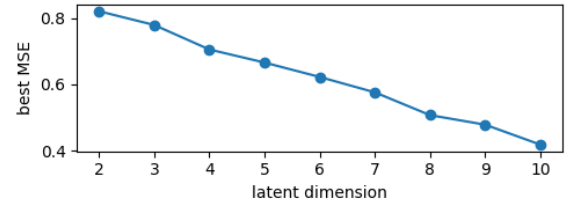


Figure 6: Reconstruction loss as a function of latent dimension for the VAE.

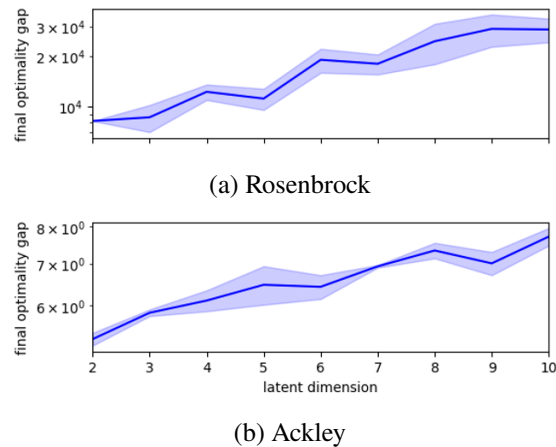


Figure 7: Latent dimension versus best BO objective value for the Rosenbrock and Ackley functions.

4.1.3 Full Tier 1 Pipeline

This work comes together in the final set of graphs for tier 1; a comprehensive pipeline of the whole process, similar to the pseudocode in Algorithm 1. Figure 8 shows results of the full pipeline applied to all four test functions below. PCA performs the best in all cases. This is expected given the nature of the input data: it contains no underlying low-dimensional structure. Therefore, PCA’s simple deterministic linear projection preserves variance without introducing distortions, producing a smooth latent representation that BO can exploit effectively. By contrast, neural networks actually perform similarly or if not worse (see Rosenbrock) than native BO. This is because this is no real manifold to uncover, so all the training the AE/VAE is doing is adding stochastic noise and distortions to the data.

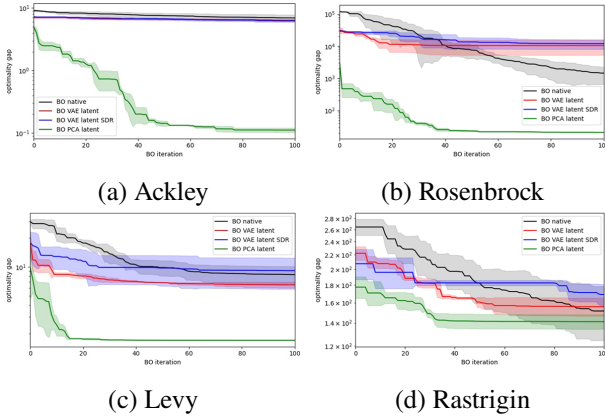


Figure 8: Tier 1 pipeline across all test functions

Overall, this tier confirms that latent space BO is only advantageous when the dimensionality reduction model uncovers a meaningful low dimensional structure. When no such structure exists, PCA serves as a good projection, whereas AE/VAEs introduce unnecessary nonlinearity that ultimately harms BO.

4.2 Tier 2

4.2.1 LVM + BO

Figure 9 shows the same pipelines ran on the same synthetic functions which now have a low-rank nonlinear underlying structure. VAEs now perform the best without exception, due to PCA struggling with nonlinearities and the VAEs ability to capture nonlinear structures. However, the additional SDR implementation only performs better in one instance, since we used the same algorithm for each function, without tuning to the specific landscape.

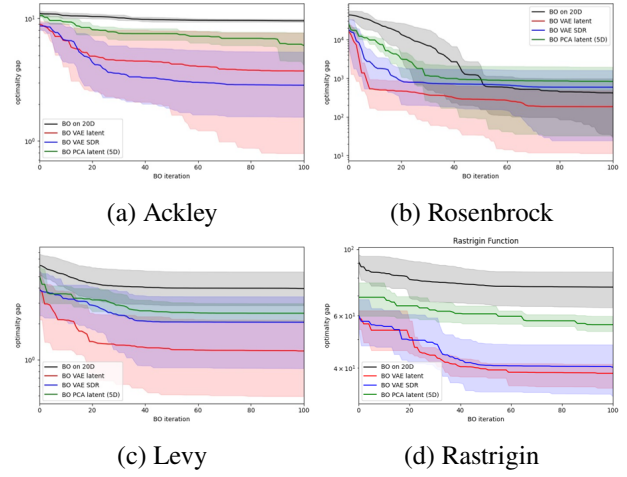


Figure 9: Tier 2 pipeline across all test functions.

4.2.2 Latent Dimensions

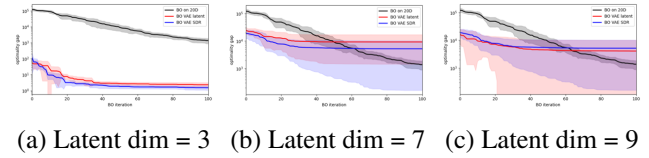


Figure 10: Effect of latent dimensionality on BO performance for the Rosenbrock function

Figure 10 and Table 3 confirms the same trend shown in Figure 7 when the data has structure. When the latent dimension is smaller than or equal to the underlying, the neural networks work to produce an optimum far below native BO. However, when the latent dimension is above (Figure 10b and Figure 10c), performance degrades heavily, due to the introduction of redundant dimensions. These mean the VAE fails to make a meaningful manifold, so the latent space is harder for the GP to model when compared to the true function.

Table 3: Best objective values obtained when running BO on the true 5D functions versus their 5D latent-space representations after dimensionality reduction.

Function	5D native	5D latent space
Rosenbrock	4.15	13.39
Rastrigin	23.97	24.68
Levy	0.45	5.67
Ackley	2.82	7.51

4.3 Tier 3 + Overall

Figure 11 shows how the VAE training has progressed over the first two tiers:

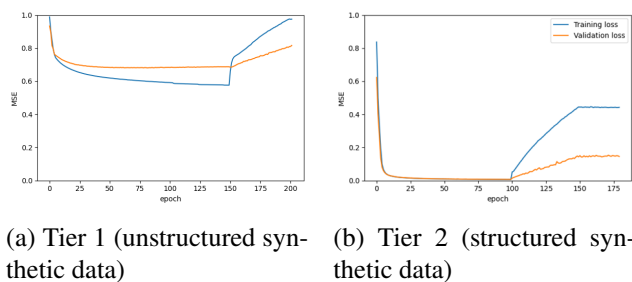


Figure 11: VAE training loss (MSE) versus epoch across tier 1 and tier 2 experiments.

Figure 11 shows the training and validation loss of the final tuned VAE at tier 1 and 2, revealing final losses of 0.665 and 0.158 respectively. This improvement in loss in tier 2 reflects the presence of a low-dimensional manifold within the structured data, which the VAE is able to learn. This is unlike the tier 1 data, confirming that meaningful manifolds are essential for effective latent space modelling.

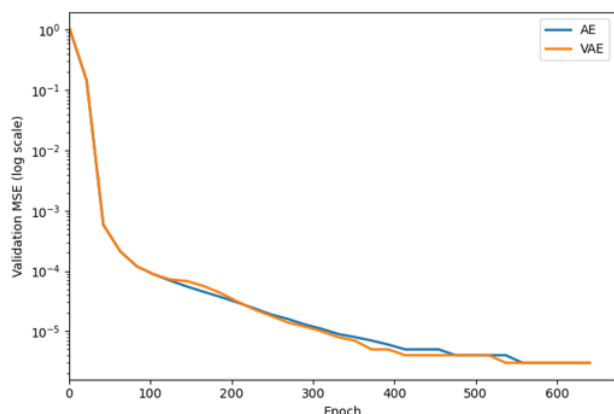


Figure 12: AE and VAE training curves for Tier 3 (reactor data).

Here, an entirely different learning profile is observed. Both LVMs quickly learn the latent space, reducing reconstruction loss by 5 orders of magnitude, down to a similar AE/VAE loss to roughly 3×10^{-6} . This decrease is expected; there is a presence of an underlying 3D physical manifold (T, tau, ruta) despite the true input being 36D. Additionally, the latent dimension was chosen to match the intrinsic dimension, meaning the VAE is not penalised by KL loss, hence its loss is near identical to the AE.

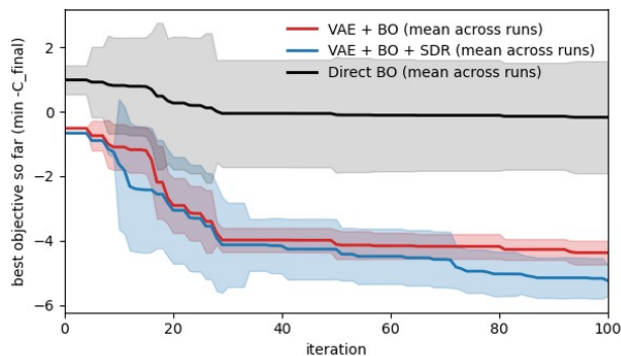


Figure 13: Results showing BO native (black), BO+VAE (red) and full pipeline (blue) on minimising negative yield of reactor

Figure 13 shows the results for tier 3 testing. Note that the yield values appear very small because the Arrhenius-controlled kinetics produce miniscule conversions under most operating conditions from the reactor constraints, giving final C concentrations in the order of 10^{-7} .

- Direct BO: Plateaus early around 0.5×10^{-7} , showing minimal improvement over the 100 iteration budget established prior. In addition, the standard variation bands show fluctuations of roughly 1.5, attesting to the inconsistencies that native BO encounters in high dimensions
- BO + VAE: Improves on native BO by reaching -4×10^{-7} , with a tighter variation of around 0.5, demonstrating the learned latent space makes the problem more optimisable
- Full pipeline: Best performance of roughly -4.5×10^{-7} , with the lowest variability of 0.25. SDR is also seen to accelerate early progress through eliminating unproductive regions and focussing on promising regions.

The tenfold improvement obtained from switching from native BO to the full novel pipeline (BO+VAE+SDR) confirm the core claim of the paper: latent modelling with SDR yield the most sample-efficient and consistent optimisation.

5 Conclusion and Outlook

Across all 3 tiers, the initial question posed by this paper was answered; BO with a latent representation is firmly superior to running BO in a higher dimensional space. Any LVM allows the surrogate to see a smoother space, which translates into a better objective. The type of LVM which is most suitable is nuanced; tier 1 shows PCA is most suitable due to its simple nature on more linear problems, and VAE/AEs are overqualified for the task, introducing stochasticity through their

training. The narrative flips with tier 2 when a curved low rank manifold is to be discovered, which more accurately represents chemical engineering problems. Hence, VAEs are pursued in tier 3, again, outperforming native BO, and when used in conjunction with SDR, the highest average yield with the tightest confidence bands is obtained. Additional findings include the complex relationships between reconstruction loss, latent dimension and objective, VAEs outperform AEs despite having worse losses, with a similar story observed from low latent dimensions with higher losses giving better objectives as well.

The most obvious avenue for future work is to apply the tier 3 pipeline to more challenging, true to life chemical engineering problem, such as coil geometry optimisation. Such work couldn't be completed in this paper due to computational and time budgets. The study could also explore richer latent models with various loss functions, such as cross entropy loss, in addition to researching alternative surrogates and kernels. Finally, the most practical opportunity which could arise from this study would be to package the tier 3 workflow such that process engineers can plug in existing simulations or code bases to obtain latent representations, where they can find an optimum to their problem. They can have the option to pick their LVM depending on the linearity and complexity of their problem, which may move theoretical ideas such as BO+VAE+SDR into a usable component in day to day optimisation.

6 Acknowledgements

We would like to thank Mathias Neufang for his support and guidance throughout this project.

7 Supplementary Information

The code for the implementation can be found in the following GitHub repository: <https://github.com/kathiir17/FYR>

References

- [1] C. J. Taylor. A brief introduction to chemical reaction optimization. *Chemical Reviews*, 2023.
- [2] C. K. Harris, D. Roekaerts, and A. K. Kuczaj. Computational fluid dynamics for chemical reactor engineering. *Chemical Engineering Science*, 1996.
- [3] Andreas Griewank and Andrea Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, 2008.
- [4] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2 edition, 2006.
- [5] Ziyu Wang, Frank Hutter, Masrour Zoghi, David Matheson, and Nando de Freitas. Bayesian optimization in a billion dimensions via random embeddings. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1778–1784, 2013.
- [6] David Eriksson, David Bindel, and Andrew Gordon Wilson. Scalable global optimization via local bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 34, pages 833–845, 2021.
- [7] Yanan Sui, Alkis Gotovos, Joel W. Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 997–1005, 2015.
- [8] Alexander Mitsos, Benoît Chachuat, and Paul I. Barton. Machine learning and optimization in chemical engineering. *Computers & Chemical Engineering*, 113:153–165, 2018.
- [9] Ian T. Jolliffe. *Principal Component Analysis*. Springer, 2002.
- [10] Geoffrey E. Hinton and Ruslan R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.
- [11] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2014.
- [12] Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 4(2):268–276, 2018.
- [13] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. MIT Press, 2006.
- [14] Emmanuel A. Chanona, Susumu Sakata, Qing Chen, Lewis Deakin, Sowmya Balaji, Dongda Zhang, and Nilay Shah. A review of derivative-free optimization methods in chemical engineering. *Computers & Chemical Engineering*, 2022.
- [15] Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492, 1998.
- [16] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.
- [17] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pages 1015–1022, 2010.
- [18] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 1278–1286, 2014.
- [19] Stephen K. N. Portillo, John K. Parejko, Jorge R. Vergara, and Andrew J. Connolly. Dimensionality reduction of sdss spectra with variational autoencoders. *The Astronomical Journal*, 160(45):1–17, 2020.
- [20] Mathias Neufang, Emma Pajak, Damien van de Berg, Ye Seol Lee, and Ehecatl Antonio del Rio Chanona. Surrogate-based optimization techniques for process systems engineering. Imperial College London teaching material, 2023. Available upon request; internal course resource.
- [21] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631. ACM, 2019.
- [22] Zavalab. Vae reactor notebook. https://github.com/zavalab/ML/blob/master/bayesianvae/Reactor/VAE_Reactor.ipynb. Accessed: 2025-02-05.

Application and Optimisation of Symbolic Regression in Crystallisation Modelling

Mingze Chen and Zhengxi Cheng

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Crystallisation is an important step in peptide drug manufacturing, but is costly and time-consuming to study through large-scale experiments, often yielding sparse data and poorly identifiable, high-dimensional kinetic parameters. This study researched a hybrid machine learning framework that combines neural ordinary differential equations (NODEs) networks with a symbolic regression (SR) model to recover both predictive and interpretable crystallisation kinetics from limited training resources. NODE is trained on only three simulated trajectories and used to generate derivative data across varying initial concentrations as samples, while the following SR searches and provide explicit augmented ODEs governing the system. In terms of system capability, the reconstructed dynamics and kinetic expressions were able to accurately match the Method of Moments experimental behaviour within ± 1 – 2 standardised concentration units of the training initial conditions, and a wider spacing of training concentrations significantly extended the effective prediction range. For the system optimality, sensitivity analyses show that 1-2 augmented dimensions under a complexity level constraint of 14-16 would yield reasonably good results for both SR prediction and NODE training. This is a result of the trade-off between interpretability, data burden and accuracy. The approach provides a data-efficient pathway for interpretable kinetic modelling of crystallisation systems and insights for future mechanistic-based design. However, its validity remains inherently local and dependent on model complexity, dimensionality, and training-set design.

1. Introduction

Peptide therapeutics have a dominant influence in metabolic diseases, accounting for 32.2% of the global market share. Specifically in diabetes, GLP-1-class peptides such as semaglutide, which support a global peptide-drug market projected to grow from \$41.44 billion in 2023 to \$68.83 billion by 2028, attract constantly increasing investment from the pharmaceutical and biomanufacturing industry¹. The quality of the drug is affected by terminal separation techniques, which usually involve chromatography and crystallisation. Crystallisation, with a more cost-effective structure, higher flexibility for scale-up and a more stable thermodynamic nature, attracts broad investigation. Experiment-driven research imposes significant time and expense requirements and is poorly adaptable to reaction conditions and protein types; therefore, researchers have shifted their attention to modelling the crystallisation process.

Similar to generic reactor designs, mechanistic models have been developed to predict the crystallisation trajectory using classical nucleation theory and a single growth rule. Though they can provide valid results, this method is still suffering from source experiment scarcity, resulting in uncertain parameters and unidentifiability between parameters during modelling. The lack of

experiments left greater opportunities for machine learning (ML) to build up. The development of neural ordinary differential equations networks enables more precise parameter screening with high robustness across a large number of inputs. For example, Bradley et al. showed that in the study of styrene dehydrogenation, the NODE-based two-stage method accurately recovered strongly nonlinear kinetic parameters from only six experimental runs and with up to 10% added noise. In contrast, the conventional direct estimation approach failed to converge under the same conditions. However, all ML models operate as a ‘black box’: the system dynamics, or the governing effects of kinetic parameters, remain uninterpretable. To complement this approach, a newly rising algorithm, the symbolic regression (SR), was employed in this article to capture underlying kinetics.

The SR was first introduced in 1992 by John R. Koza in his *Genetic Programming*. Rather than embedding the system within a black box complex, SR searches for targeted parameters and functional patterns directly from data, through its ‘gene-like’ inheritance network, yielding the best-matching equations between inputs and outputs. Its high interpretability and wide adaptability enable the

establishment of a model without much prior knowledge, the verification of physical formulae, and the completion of partially known models. It has been massively applied in electrical engineering and material science. In this research, an SR model will be constructed upon a NODE model, aiming to co-interpret images of dynamic crystallisation systems.

2. Background

From mechanistic descriptions to industrial process design, the kinetic behaviour of crystallisation has been extensively investigated. In protein crystallisation, studies span multiple scales—from molecular-level nucleation mechanisms to the development of population balance models (PBMs)². These models employ differential equations to describe the changes in key physical metrics, such as particle size distribution, crystal morphology, filterability, flowability, and stability, as well as their spatial and temporal dynamics. Crystals with various facets and habits are also considered, with the PBM model being extended to account for morphology-dependent kinetics³. The development of these models and tests enabled engineers to determine operational features such as the acceptable supersaturation range, the metastable zone width, nucleation-dominated and growth-dominated regimes, as well as to predict optimal cooling curves for batch processes. However, aspects such as crystalliser stage design, attrition and breakage kinetics, and scale-up strategies still heavily rely on empirical formulae and heuristic rules^{4,5}. It triggers interest in research into the application of machine learning applications, which potentially could perform more time-saving screenings and simulations and accelerate the design of early-stage predictive tools.

Thanks to their strong transferability, powerful pattern-recognition capability, and the widespread success of deep learning in visual tasks, most machine learning models in crystallisation research have been used in problems such as inferring crystal habits and images, classifying crystallisation types, assisting solvent selection, predicting solubility, and reducing experimental workload. In these applications, ML models primarily operate more as decision tools, giving a “yes-or-no” judgement or providing categorical suggestions. By contrast, far fewer studies have attempted to use machine learning to uncover the underlying mechanistic rules or to model crystallisation kinetics, leaving this aspect of the field comparatively new. A neural network (NN)-based solution has been proposed to address complex kinetics, uncertain parameters, and insufficient experimental data, providing better fits to a conventional mechanistic model. It also demonstrates optimistic transferability by making reasonably accurate predictions when faced with 3

new, different templated-induced nucleation systems without prior learning experience. The model's strong performance has prompted researchers to explore the limits of its ability to transfer knowledge and generalise beyond its training domain. It also raises a bigger question: to what extent can the mechanisms of crystallisation be expressed through design equations or other forms of generalised kinetic descriptions? This motivation led to an examination of machine-learning tools capable of revealing such structures, among which symbolic regression emerged as a particularly promising candidate. The study, therefore, builds upon its core capability, the discovery of explicit mathematical expressions, to investigate whether crystallisation behaviour can be represented in a more generic mechanistic form, like a well-studied reactor design.

3. Methodology

A complete simulation cycle consists of three key stages: sampling, SR training, ODE solving and visualisation. The behaviour of crystallisation was analysed with respect to two quantities over time t : concentration (denoted c , or x_0 in coding) and particle size (denoted d_{43} , or x_1 in coding). The full training loop is Python-based, while the symbolic regressor package in the Julia environment was invoked during training.

3.1 Sampling

3.1.1 The Mechanistic Model

The study focuses on how symbolic regression solves underlying dynamic relationships within a process system, where both measurable readings of decisive variables and their time derivatives are required. The former reveals the overall states of the system, while the latter provides instantaneous information on how the system evolves in time as a vector field. The sampling section consists of a mechanistic model of crystallisation followed by a NODE network, aiming to generate sufficient sample data for further SR training⁶. The given mechanistic model conducted six crystallisation experiments *in silico* with different initial conditions: 3 classified in the ‘Train Set’ and the other 3 in the ‘Test Set’, each providing both concentration- and particle-size-time profiles, serving as the ground-truth dynamite of this system. Internally, these experiments are evaluated based on the classical nucleation theory (Solve of Method of Moments ODEs, **MoM**), and a single-nucleus growth correlation, and these resulting points on solution curves are processed with significant noise and discontinuity, hindering accurate derivative evaluation; thus, a NODE model is applied to help collect time-derivative information.

3.1.2 The role of NODE

Like any conventional machine learning model, NNs learn relationships between inputs and outputs using a human-brain-like pattern. Input information is passed through layers of neurons in a network, each acting as a linear regression model with its own weight and bias, which capture the influence of each input and reveal outputs. Extending this idea further, a NODE has been designed to express real-time systematic variations by modelling its ordinary time differential equations by NNs as below,

$$\frac{dz(t)}{dt} = f_{\theta}(z(t), t) \quad (Eq\ 3.1.1)$$

Where z is the latent state vector at time t ; f_{θ} is a neural network parameterised by θ , representing the right-hand side of the learned dynamics.

By putting an NN on the right-hand side of an ODE, it is trained to predict time derivatives as outputs after acquiring mechanistic datasets. The implementation of the NODE can be divided into two steps: training the model and performing sampling.

3.1.3 Training the NODE

The initial conditions of mechanistic experiments from the ‘Train Set’ are directly fed into the NN to predict time derivatives. An affiliated internal ODE solver will then integrate these derivatives to give the fresh ‘learned trajectories’, which are compared with the mechanistic trajectories previously generated, and their absolute deviation is quantified as a loss function, which is used as a guidance for returned NN parameter adjustment. It is also the main quantity that other ML models use to quantify the accuracy of prediction.

$$L = \frac{1}{N} \sum (\hat{y}_i - y_i)^2 \quad (Eq\ 3.1.2)$$

Where L denotes the loss function, y_i denotes the reference or true value at sample i , $\hat{y}_i$ is the model-predicted value at the same sample, and N is the total number of samples used in the evaluation.

Meanwhile, the other three mechanistic trajectories within the ‘Test Set’ remain unseen by the NODE. The ‘Test Set’ serves as a quality gate of the NODE to assess its generalisation ability, the predicting performance on new initial conditions, which work as alarms to any necessary manual parameter adjustment when the model fails on unseen states. Besides, it is vital to notice that two augmented quantities, a_0, a_1 , are added to the training pool with c, d_{43} , which will eventually be found, that will help achieve more accurate NODE

fitting. Detailed discussions in Section 4.5. Once a thoroughly tested number of iterations, 1200, on the NODE is reached, the training stops. Eventually, the NODE has formed its own predictive trajectories with high fidelity towards the mechanistic ones, erasing the discreteness and discontinuity from the raw trajectories, and can provide valid training sources for the subsequent SR model establishment.

3.1.4 Sampling by the NODE

In this section, the ‘Train’ and ‘Test’ identities of all six experiments are removed, and their initial conditions are reshown to the fully trained NODE. For each state-time pair along the mechanistic trajectories supplied to the NODE, their time derivatives are independently evaluated in the output domain, as shown in the schematic shown below⁷.

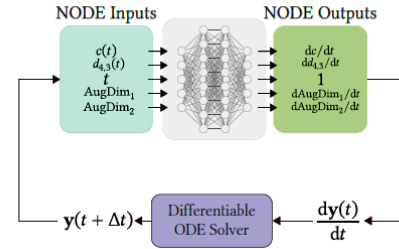


Figure 3.1.1 NODE working cycle

This procedure creates two sampling datasheets as listed below,

An ‘all-points’ datasheet containing 50 uniformly sampled state-time pairs from each of the six mechanistic experiments. Each sampled point includes the mechanistic state variables, the augmented variables, and the instantaneous time derivatives provided by the trained NODE.

experiment_id	train_id	t	TYPE	x0	x1	a0	a1	dc/dt	ds/dt	da0/dt	da1/dt
0	0	0	TRAIN	0.483247	-1	0	0	0.137891	3.8737	0.78792	-0.40986
0	1	0.020408	TRAIN	0.483823	-0.92028	0.016605	-0.00778	0.132272	3.938389	0.840915	-0.35063
0	2	0.040816	TRAIN	0.48781	-0.83929	0.03439	-0.01423	0.078878	3.997051	0.90381	-0.27983
0	3	0.061224	TRAIN	0.489013	-0.75718	0.053571	-0.01911	0.036058	4.049301	0.977813	-0.19607
0	4	0.081632	TRAIN	0.489199	-0.67413	0.073385	-0.02214	-0.02003	4.089706	1.044126	-0.09798
0	5	0.102041	TRAIN	0.488098	-0.59033	0.097096	-0.02301	-0.06037	4.119767	1.10385	0.01722
0	6	0.122449	TRAIN	0.4854	-0.50607	0.121986	-0.02138	-0.17065	4.13592	1.157331	0.140949
0	7	0.142857	TRAIN	0.480753	-0.42163	0.148552	-0.01693	-0.29158	4.135569	1.06449	0.28032
0	8	0.163265	TRAIN	0.473774	-0.3374	0.175949	-0.00929	-0.40561	4.11597	1.549333	0.457948
0	9	0.183673	TRAIN	0.464862	-0.25378	0.212677	0.001871	-0.54052	4.074397	1.705036	0.637978

Table 3.1.1 NODE dense time sampling “all-points’ datasheet” example

A Latin hypercube sampled sheet where 2048 more dense and randomly taken points in the possible space are analysed in the same way by the NODE and collected. It is acting as a supplementary or diagnostic sample to reflect the robustness, stability and physical consistency of the learning results, and it is not directly used for further SR training.

3.2 Training the SR Model

The sample is fed into the file for SR training, where there is another deep learning model: the symbolic

regressor. Similarly, it aims to find potential relationships between state and augmented variables as inputs and their time derivatives as outputs. The symbolic regressor operates in a manner comparable to genetic programming; it starts the looping with searching around pooled algebraic structures and calculation operators (i.e. +, -, ×, ÷, log) for potential mathematical equation candidates. Apart from the loss, the equation is evaluated from another angle, the complexity. The complexity of a candidate equation refers to a numerical measure of its structural richness. It measures how complicated the mathematical expression is, based on the size and composition of its underlying expression tree. The table below is the weighting scheme, used by the symbolic regressor in this research, which assigns:

Component	Complexity Contribution
Variables and constants	1
Addition and subtraction (+, -)	1
Multiplication and division (×, ÷)	2
Power operator (^)	4
Transcendental functions (exp, log, sin, etc.)	3–5

Table 3.2.1 Example of complexity contributions of different nodes.

Thus, the total complexity is given by:

$$\text{Complexity} = \sum_{i=1}^N w_i \quad (\text{Eq 3.2.1})$$

where w_i is the weight assigned to the i -th node in the expression tree and N is the total number of nodes in its passed genetic layer.

At each iteration, expressions with lower prediction error or lower structural complexity are retained, while inferior candidates are discarded, and new candidates are generated by maintaining and recombining surviving expressions, allowing the search to progressively improve the equation space. Eventually, this process yields a Pareto-optimal set of models, from which the expression with the minimal loss under fixed complexity limits is selected.

$$\begin{cases} \frac{dc}{dt} = f(c, d_{43}, a_0, a_1, t) \\ \frac{dd_{43}}{dt} = g(c, d_{43}, a_0, a_1, t) \\ \frac{da_0}{dt} = h(c, d_{43}, a_0, a_1, t) \\ \frac{da_1}{dt} = j(c, d_{43}, a_0, a_1, t) \end{cases} \quad (\text{Eq 3.2.2})$$

3.3 ODE solving and visualisation

After the four derivatives were properly expressed in human-readable equations, the final stage was to validate model performance. Obtained ODEs regarding x_0 and x_1 are solved along the standardised time scale [0,1] by a standard ODE solver, producing their concentration and particle size trajectories, named as Predicted lines. The six initial conditions are imported again into the mechanistic model, and the resulting mechanistic trajectories are plotted on the same grid, named as Observed lines. The model's effectiveness in recovering underlying crystallisation kinetics can be roughly determined by the deviation between the two curves as follows. A detailed discussion will be given in Section 4.3.

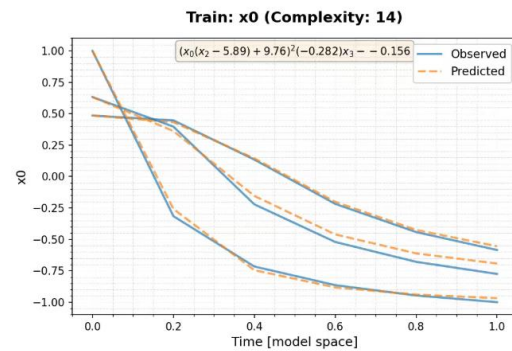


Figure 3.3.1 Result of solving the ODE of $\frac{dc}{dt}$ at 3 Training Initial Concentrations.

4. Results & Discussion

4.1 Trade-off between ODE Complexity and Loss

Symbolic regression typically produces a family of candidate expressions with complexities ranging from 1 to the self-defined maximum allowable complexity (30 in this case). For each complexity level, the corresponding loss value is computed as well. These two key quantities construct intuitive criteria for evaluating analytical expressions:

1. **Lower complexity is desirable**, as excessive structural complexity reduces interpretability.
2. **Lower loss is also desirable**, as it reflects greater fidelity to the underlying data.

Under a fixed set of training samples, the empirical relationship between the equation complexity and loss exhibits the trend shown in **Fig.4.1.1** and

Fig.4.1.2.

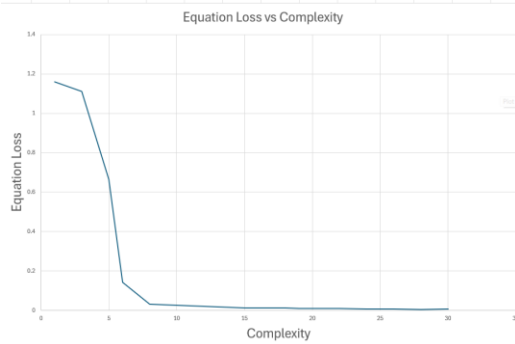


Figure 4.1.1 Equation Loss vs Complexity of ODE $\frac{dc}{dt}$.

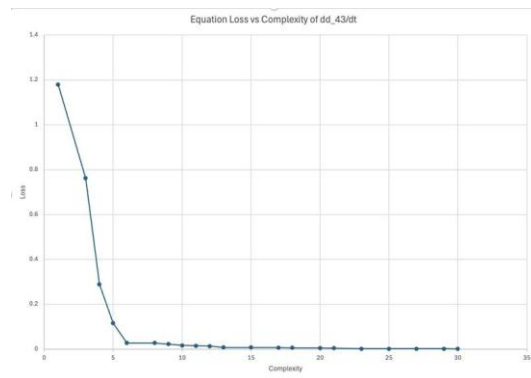


Figure 4.1.2 Equation Loss vs Complexity of ODE $\frac{dd_{43}}{dt}$.

While loss generally decreases with increasing complexity, the rate of improvement witnesses a plateau. In other words, once the complexity becomes sufficiently large, further increases yield only marginal reductions in prediction error.

This behaviour enables a principled trade-off: within a specific complexity range, the model achieves sufficiently accurate predictions without incurring extra unnecessary structural complexity. As illustrated in both **Fig. 4.1.1** and **Fig. 4.1.2**, the loss function becomes nearly invariant beyond a complexity of approximately 14. Therefore, during equation selection, a complexity threshold of **14** is adopted as the most feasible choice, with no apparent need to extend the threshold to higher values such as 17.

4.2 Sensitivity Analysis

The preceding validation establishes that loss decreases with increasing complexity. To investigate the behavioural characteristics of the SR system and identify factors that influence its performance, two sensitivity tests were conducted while keeping the complexity fixed:

1. At a fixed complexity of 16 and fixed dimensionality, the number of experimental groups was increased.
2. At fixed complexity and fixed 6 of experimental groups, the system dimensionality was increased.

The results are presented:

Number of Experiment Groups	Average Loss of the ODE system
6	0.023
7	0.037
8	0.099

Table 4.2.1 Loss vs. number of experimental groups

Total Dimensions	Average Loss of ODE system
3 (c, d_{43}, t)	0.160
4 (c, d_{43}, a_0, t)	0.032
5 (c, d_{43}, a_0, a_1, t)	0.029
6 ($c, d_{43}, a_0, a_1, a_2, t$)	0.065

Table 4.2.2 Loss vs. dimension

Table 4.2.1 clearly shows that, under the same equation complexity, the loss increases as the number of experimental groups increases. This observation aligns with intuition: as more experiments are added, the information burden that a fixed-complexity expression must simultaneously capture increases. A model of unchanged structural complexity is required to accommodate a larger and more diverse dataset; consequently, its precision on each individual dataset naturally declines, leading to an increase in overall loss.

As Table 4.2.2 shows, increasing the dimensionality effectively increases the flexibility of the ODE system. With more available variables, the SR model gains additional degrees of freedom to describe the same collection of experimental data. From this perspective, one would expect that—for fixed complexity—a higher-dimensional system should reproduce the data more accurately, resulting in lower loss.

However, the actual results deviate from this expectation: as shown in the second table, the average loss of the complete ODE system exhibits a local minimum at total dimension = 4. This phenomenon is interpreted as follows. Increasing dimensionality indeed increases the expressive flexibility of the ODE system; under fixed complexity, this tends to reduce the loss. However, adding dimensions also increases the total

amount of data (each augmented dimension introduces an additional input column). According to the reasoning from Table A, an increase in total data requires the model to capture more information without increasing complexity, which drives the loss upward.

The two opposing effects—greater flexibility, lowering loss, and greater data volume raising loss—interact to produce the observed local minimum when the total dimensionality reaches 4.

Summary of SR characteristics and behaviour

In summary, the behaviour of symbolic regression in reconstructing the ODE system can be characterised as follows:

1. A more complex ODE system leads to a more accurate description of the data, resulting in a lower loss.
2. For ODE systems of the same structural complexity, increasing the amount of training data makes the SR fit less precise, thereby increasing the loss.
3. For ODE systems of the same complexity, increasing the dimensionality generally improves the accuracy of SR, as the system gains greater expressive flexibility.
4. However, increasing dimensionality inevitably increases the total data volume, which tends to worsen the SR fit.

Due to the competition between these two opposing effects—greater flexibility (improving accuracy) and greater data volume (reducing accuracy)—the loss reaches its minimum when the total dimensionality equals four.

4.3 Using NODE to Reconstruct Dynamical Systems

NODE learns the c - t and d_{43} - t trajectories from three internally generated simulation experiments.

These three experiments were intentionally chosen to mimic a situation with limited experimental data. After training, the NODE can generate the trajectories of $c(t)$ and $d_{43}(t)$, as well as their corresponding derivatives.

In this study, the simulated experimental trajectories were themselves obtained by solving a traditional crystallisation model: the Method of Moments (MoM) ODE system. Noise was added to simulate real experimental situations of an unknown

system. When the same initial concentrations were used, the trajectories produced by solving the augmented ODE system recovered from NODE+SR(Predicted) match the original MoM trajectories (Observed) very well (as shown in Fig.4.3.1).

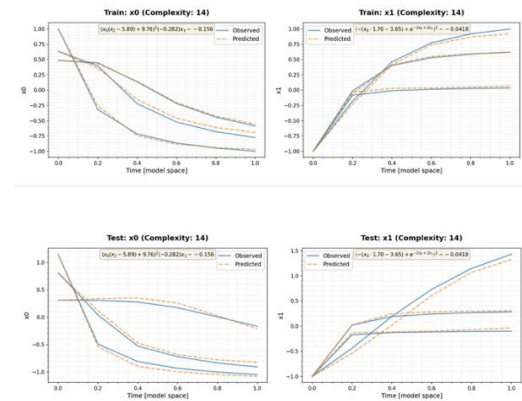


Figure 4.3.1 Comparison of results by solving the ODE of $\frac{dc}{dt}$, $\frac{dd_{43}}{dt}$ at 3 Training ICs (top two graphs), and at 3 Testing ICs.

This highlights the significance of the augmented ODE system: when facing an unfamiliar physical system, one may perform only a small number of experiments, train a NODE on these few trajectories, and then use SR to recover a corresponding dynamical equation set. Although the augmented dimensions do not always correspond to physically interpretable quantities, the learned ODE system can still be solved for any initial condition to obtain crystallisation trajectories. This offers a potentially large reduction in experimental cost and time.

4.4 Limitations

However, such a system inevitably has its limitations.

Two tests were conducted in this study.

4.4.1 Testing NODE predictions across a wide range of initial concentrations

After NODE was trained on three trajectories corresponding to initial concentrations of 15.5, 16.5 and 19 (3 red points), its predictive capability was evaluated across 26 unseen initial concentrations ranging from 5 to 30 (26 blue points). Prediction accuracy was quantified by computing the RMSE between the NODE-generated trajectories and the simulated experimental trajectories.

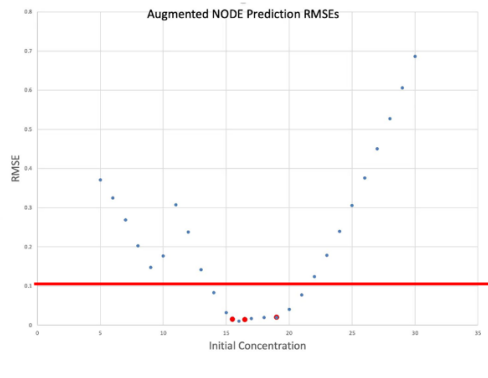


Figure. 4.4.1 RMSE of NODE predictions at different ICs (5-30)

The results (Fig. 4.4.1) indicate that NODE achieves an RMSE below 0.1 only within the approximate range:

$$c_0 \in [13.5, 21.2],$$

which corresponds to roughly ± 2 concentration units from the boundaries of the training set. This outcome suggests that the NODE model yields reliable predictions only within a narrow neighbourhood of the training domain.

4.4.2 Effect of increasing spacing between training trajectories

To further examine the generalisation limitations of NODE, an additional experiment was conducted in which the number of training trajectories remained fixed, but the spacing between them was deliberately increased. Instead of using closely spaced initial concentrations, the training concentrations were set to 7, 16.5, and 25. The objective was to determine whether wider separation between training conditions would broaden the range over which NODE can make accurate predictions.

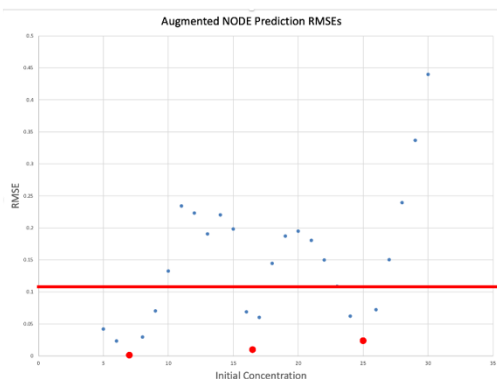


Figure. 4.4.2 RMSE of NODE (Trained at 7, 16.5, 25) predictions at different ICs

The results (Fig.4.4.2) reveal that each training concentration produces its own “region of

influence,” within which the NODE-generated trajectories exhibit acceptable accuracy. In all cases, this region extends approximately ± 2 concentration units from the training point. When the training concentrations are widely separated, these non-overlapping regions collectively cover a much broader portion of the concentration space—even though the total number of training trajectories remains the same.

This observation suggests a practical strategy for improving generalisation under limited data availability: carefully spacing the training conditions can substantially expand the effective predictive range of the NODE model, without increasing the amount of training data. By ensuring that the influence regions do not overlap, the overall usable coverage can be significantly enlarged while preserving the model's accuracy within each local neighbourhood.

4.5 Investigation of the Role of Augmented Dimensions in NODE

To evaluate the functional contribution of the augmented dimensions in the NODE system, NODE was tested under four settings with augmented dimensions of 0, 1, 2, and 3. Training trajectories were generated using initial concentrations of $c_0 = 15.5, 16.5, 19$, and testing trajectories were generated using $c_0 = 14.3, 17.7, 20$. For each setting, the average training RMSE and testing RMSE were computed, and the results are summarised in Table 4.5.1.

Number of AugDims	Mean RMSE for Training	Mean RMSE for Testing
0	0.2949	0.3946
1	0.0133	0.1579
2	0.0114	0.1842
3	0.0118	0.1918

Table 4.5.1 NODE prediction RMSEs when trained with different dimensionalities.

Table 4.5.1 indicates that NODE predictions are barely accurate if 0 Augmented Dimensions are allowed while training. A conclusion can be drawn that augmented dimensions are able to assist the NODE in making predictions more accurately. The training RMSE reaches a local minimum when the augmented dimension equals 2, whereas the testing RMSE reaches a local minimum when the augmented dimension equals 1.

This behaviour mirrors the earlier observation from SR dimensionality analysis: in SR, a total ODE dimensionality of four (corresponding to two augmented dimensions) yielded the lowest loss, and in NODE, one or two augmented dimensions yield

the lowest RMSE for training and testing, respectively.

5. Conclusion

This work investigates a hybrid NODE + SR framework for recovering the governing dynamics of a crystallisation process from limited data. Symbolic regression is shown to follow clear behavioural rules: for a fixed equation structure, increasing data quantity raises the loss, while increasing dimensionality initially improves accuracy but eventually worsens it due to increased data load. The SR loss–complexity relationship reveals a diminishing-return trend, enabling the selection of an optimal complexity threshold (≈ 14) that balances predictive accuracy and interpretability.

After training on only three simulated trajectories, NODE can generate derivative data for arbitrary initial concentrations, from which SR reconstructs an augmented ODE system. This system successfully replicates the original MoM dynamics within ± 1 – 2 concentration units of the training domain but fails to generalise beyond it. Sensitivity analyses of augmented dimensions show that one to two augmented dimensions provide the best performance for both NODE fitting and SR equation accuracy.

Further investigation into NODE’s generalisation behaviour reveals that each training concentration induces a local region of influence of approximately ± 2 concentration units. When the training concentrations are deliberately spaced farther apart, these non-overlapping influence regions collectively cover a much broader segment of the concentration space, even though the total number of training trajectories remains unchanged. This indicates that careful selection and spacing of training conditions can substantially expand the effective predictive range of the NODE model under limited data availability.

Overall, the NODE + SR approach offers an efficient method for extracting interpretable dynamical equations from sparse experiments, but its validity is inherently local and strongly dependent on model complexity, dimensionality, data coverage, and the experimental design of training conditions.

6. Outlooks

Future work will extend the present algorithm to more complicated and practical crystallisation systems beyond the classical nucleation theory and the growth mode adopted in this study. Industrial crystallisation frequently involves secondary nucleation mechanisms, size-dependent growth behaviour, agglomeration–breakage pattern, and seeded crystallisation processes, all of which

introduce richer kinetics than those in the current simple model. Modelling such mechanisms would push the NODE to learn higher-dimensional and more realistic vector fields. At the same time, symbolic regression could be used to express the process of more substances than protein or peptide, even adapting to different facets and habits. Recovering more readable differential forms that capture these advanced behaviours. Further extensions may also consider non-classical nucleation pathways, polymorphic transitions, or temperature-dependent crystallisation, providing a broader context in which to evaluate the generality and robustness of the NODE–SR framework.

Besides, a useful direction for future work is to interpret the discovered differential equations in terms of the underlying crystallisation physics. The symbolic expressions obtained from SR encode the effective kinetic relationships learned from data, but their coefficients and functional dependencies require systematic physical interpretation. For example, terms involving supersaturation-like quantities may correspond to primary nucleation or surface-integration-controlled growth. At the same time, dependence on crystal size or concentration may indicate size-dependent kinetics or mass-transfer limitations⁸. Similarly, nonlinear combinations of state variables could reveal interactions not explicitly included in the mechanistic model, such as latent secondary nucleation pathways or growth inhibition effects. By mapping each term in the recovered ODEs back to plausible physical mechanisms, the framework can serve not only as a model-identification tool but also as a means to generate new mechanistic hypotheses and refine existing crystallisation theory.

Another promising direction is to incorporate transfer learning⁷ to integrate existing mechanistic knowledge into the model-development pipeline. In the current framework, the Neural ODE is trained from scratch for each crystallisation scenario, even though many crystallisation systems share common structural behaviours—such as supersaturation-driven growth, nucleation trends, or size-dependent kinetics.

By reusing portions of a previously trained NODE as an informed initialisation, or by selectively fine-tuning specific layers on new datasets, the learning process could be significantly accelerated while preserving physically meaningful patterns already captured in earlier models. Such a transfer-learning strategy would not only reduce computational cost but would also enable the framework to adapt more efficiently when extended to new solutes, operating regimes, or mechanistic models, ultimately strengthening the generality and usability of the NODE–SR methodology.

7. Acknowledgements

The authors would like to thank their supervisors, Daniele Pessina and Maria Papathanasiou, for sharing their deep technical expertise in machine learning modelling and coding modification, and for their patient and insightful supervision and guidance throughout this meaningful 3-month period.

8. References

1. Wani, G. and Faizullabhoj, M. (2025). *Peptide Therapeutics Market - By Type, By Application, By Route of Administration, By Manufacturer Type, By Synthesis Technology, By End Use - Global Forecast, 2025 - 2034*. [online] Global Market Insights Inc. Available at: <https://www.gminsights.com/industry-analysis/peptide-therapeutics-market> [Accessed 10 Dec. 2025].
2. Liu, J.J., Ma, C.Y., Hu, Y.D. and Wang, X.Z. (2010). Modelling protein crystallisation using morphological population balance models. *Chemical Engineering Research and Design*, 88(4), pp.437–446. doi:<https://doi.org/10.1016/j.cherd.2009.08.015>.
3. Ma, C.Y. and Wang, X.Z. (2012). Model identification of crystal facet growth kinetics in morphological population balance modeling of l-glutamic acid crystallization and experimental validation. *Chemical Engineering Science*, 70, pp.22–30. doi:<https://doi.org/10.1016/j.ces.2011.05.042>.
4. Bermingham, S.K., Neumann, A.M., Kramer, H.J.M., Verheijen, P.J.T., van Rosmalen, G.M. and Grievink, J. (2000)
5. Ochi, T., Bolanos-Garcia, V.M., Stojanoff, V. and Moreno, A. (2009). Perspectives on protein crystallisation. *Progress in Biophysics and Molecular Biology*, 101(1-3), pp.56–63. doi:<https://doi.org/10.1016/j.pbiomolbio.2009.12.001>.
6. Pessina, D., Calderon De Anda, J., Heffernan, C., Heng, J.Y.Y. and Papathanasiou, M.M. (2025). Integrated In Vitro/In Silico Uncertainty Quantification Method for Protein Crystallization Models. *Industrial & Engineering Chemistry Research*, 64(24), pp.12025–12035. doi:<https://doi.org/10.1021/acs.iecr.4c04517>.
7. Pessina, D., Tian, T., Watson, O., Heng, J.Y.Y. and Papathanasiou, M.M. (2025). Transfer Learning Assessment of Data-driven Crystallisation Processes via Constrained Neural Ordinary Differential Equations. *bioRxiv* (Cold Spring Harbor Laboratory). doi:<https://doi.org/10.1101/2025.09.22.67773>.
8. Fedorov, A., Perehodjuk, A. and Linke, D. (2023). Kinetics-constrained neural ordinary differential equations: Artificial neural network models tailored for small data to boost kinetic model development. *Chemical Engineering Journal*, 477, p.146869. doi:<https://doi.org/10.1016/j.cej.2023.146869>.

Lignin-Directed Pore-Mouth Engineering Enables Tunable Sodium Storage in Hard Carbon Anodes

Liguang Zhu, Tong Wu

Department of Chemical Engineering, Imperial College London, U.K.

Abstract: The pore architecture of hard carbon plays a critical role in sodium-ion storage, and tailoring the pore mouths—rather than the pore bodies—offers a more effective and controllable design pathway. In this work, we introduce a pore-mouth engineering strategy in which lignin is employed as a regulating agent to modify the entrances of micropores in commercial YP-50F. Lignin physically adsorbs onto the carbon surface and contracts during carbonisation, forming an in situ coating layer around the pore mouths, thereby reducing the accessible surface area and generating additional sub-nanometre pores. Systematic comparisons across carbonisation temperature, lignin/AC ratio, and pre-oxidized lignin–AC hybrids demonstrate that pore-mouth tailoring markedly influences the initial Coulombic efficiency (from ~15% to ~70%), plateau behaviour, and cycling stability. The optimized sample achieves a reversible capacity of 185 mAh g⁻¹, confirming the functional effectiveness of tuning the pore-mouth structure. Preliminary testing of pre-oxidised lignin composites shows that, while the specific capacity is not increased, the reversibility is significantly enhanced, and stronger pore-mouth constriction is achieved. This suggests that future strategies involving selective functional-group modification of lignin may further enhance the controllability and efficiency of pore-mouth engineering. Raman, XRD, and N₂ adsorption analyses further confirm that adjustments in pore-mouth configuration directly correlate with variations in microcrystalline structure, accessible surface area, and corresponding Na-storage characteristics. These findings identify pore-mouth engineering as a promising structural design approach for the advancement of hard-carbon anodes.

Keywords: activated carbon, lignin, pore structure, green energy, sodium-ion batteries, in-situ characterization

1. Introduction

Currently, the rapid growth of electric vehicles, energy storage stations, and the uninterrupted power supply (UPS) industry has driven an explosion in global battery demand.¹⁻² With mature technology and high energy density, lithium-ion batteries (LIBs) have become the dominant system.³ Yet, their fundamental limitations remain unresolved as demand rises. The root cause lies in limited lithium reserves and their geographically concentrated distribution, which has led to a significant increase in the prices of lithium metal. These resource constraints heavily impact the large-scale LIB production chain and raise serious concerns over sustainability and supply security. While LIBs excel in electric transport and small electronics, their use in grid-scale energy storage is still restricted by lithium resources and cost barriers.⁴ Compared with LIBs, the relatively uniform global distribution of sodium resources has revived strong interest in sodium-ion batteries (SIBs) within the battery community. In addition, using Al foil as the anode current collector brings a clear cost advantage over the Cu foil commonly used for LIB anodes.⁵

Electrodes have long been one of the core directions in battery research. This work focuses on the anode side of sodium-ion batteries (SIBs). Within this system, hard carbon is recognised as the most widely used and commercially attractive anode material.⁶ Hard carbon originates from precursors rich in sp³-hybridised carbon frameworks, which inhibit the oriented growth and ordered alignment of graphitic layers during carbonisation. As a result, hard carbon cannot develop long-range-ordered,

well-crystallised graphite structures—consistent with its definition as non-graphitizable carbon.⁷⁻⁸

The key microstructural features of hard carbon include random, disordered, and curved graphitic-like layer stacking, as well as a high density of defects and micro- to nano-scale pores (micropores and nanopores). The stacking of these carbon sheets further enables the self-assembly into closed-pore structures (closed pores).⁹

Current studies show that closed pores serve as the dominant pore-filling storage domains for Na⁺ ions. In the low-potential plateau region (<0.1 V), the pore-filling and Na-cluster storage mechanisms contribute more than 60% of the total capacity, whereas crystallographic intercalation within ordered graphitic lattices accounts for only a minor fraction.¹⁰⁻¹¹ This indicates that the dominant capacity contribution arises from pore-dominated sodium storage rather than crystalline-lattice insertion. Consequently, regulating the pore-mouth size and designing the pore architecture are key strategies for improving low-potential plateau performance. However, the detailed mechanism of closed-pore formation and the definitive microstructure of hard carbon remain insufficiently clarified.¹²⁻¹⁴

Biomass-derived precursors are currently one of the most important fields in hard carbon research, owing to their resource sustainability, wide availability, green and environmentally benign nature, and low cost. A wide range of agricultural and industrial wastes—including rice husk, reed straw, fruit peels, and other lignocellulosic residues—have been used as anode precursors for

sodium-ion batteries (SIBs), delivering competitive reversible capacities.¹⁵⁻¹⁶

Lignin is a biomass constituent featuring a highly cross-linked three-dimensional aromatic network composed of hydroxyphenyl (H-), guaiacyl (G-), and syringyl (S-) units, which are formed through the dehydrogenation and radical polymerisation of p-coumaryl, coniferyl, and sinapyl alcohols. As the second most abundant biopolymer in nature, lignin is considered one of the most suitable sp^3 -rich, non-graphitizable precursors for producing hard carbon anodes.

The 3D network is primarily coupled by β -O-4, α -O-4, β -5, and 5-5 linkages, with β -O-4 as the dominant structural bond. This bonding motif contributes to graphitisation resistance, pore preservation, and alkali-metal storage capability.¹⁷ Systematic studies have shown that lignin-rich precursors exhibit higher inherent alkali-metal (especially Na^+) storage capability than hemicellulose- or pectin-rich counterparts, establishing lignin as one of the most representative and capacity-favourable biomass precursor classes for SIB anode development.¹⁸

However, existing studies also point out that lignin-derived hard carbon prepared by direct carbonisation commonly exhibits *poor interfacial stability*, limited low-potential plateau capacity (<0.1 V), and relatively low initial Coulombic efficiency (ICE, typically 50–80%), which severely limits practical large-scale production and commercialisation feasibility.¹⁹⁻²⁰

To address these limitations, intensive studies on lignin structural regulation have been proposed, including low-temperature pre-oxidation strategies to introduce carbonyl groups ($C=O$) and strengthen intermolecular cross-linking, thereby inhibiting graphitic layer rearrangement and excessive fusion of carbonaceous intermediates during high-temperature carbonisation.²¹ These effects are commonly associated with enlarged interlayer spacing (d_{002}) and altered defect and microvoid characteristics.²² In parallel, lignin molecular sieving and solvent-based fractionation strategies have also been developed to tune polarity-dependent functional-group exposure and molecular-weight distribution, collectively demonstrating that lignin functional-group composition and intrinsic molecular architecture directly determine the nanopore entrance (pore-mouth) configuration and reversible Na^+ storage plateau domains after carbonisation.²³

Lignin-derived hard carbon prepared by direct carbonization commonly suffers from high defect density, pore-mouth collapse/blockage, and poor interfacial compatibility, leading to intensive side reactions, SEI intrusion, irreversible surface Na^+ loss, and low ICE.²⁴⁻²⁵ Unlike studies that employ lignin as a bulk precursor, this work treats lignin

purely as a pore-mouth regulator and sealing agent on a commercial microporous scaffold, YP-50F, to enable entrance shrinkage and fusion/defect suppression at micropore openings (<1 nm), rather than new-pore formation. This design is intentionally driven by purely physical modifications based on solid-state mixing and pore-mouth narrowing, rather than chemical bonding or chemical activation pathways, highlighting a physical-fusion-dominant mechanism rather than a chemistry-governed approach. Here, the regulation strategy is based on physical structural modulation (powder mixing-induced physical coating/shrinkage) rather than chemical functional-group reactivity, reinforcing that the role of lignin is governed by physical transformation, not chemical causality. Recent mechanism studies emphasise that reversible Na^+ plateau capacity is primarily determined by closed-pore interior accessibility and pore-mouth structural state, and that tightened, defect-suppressed pore entrances reduce SEI ingress, improve desolvated Na^+ entry, and stabilise Na-cluster/pore-filling reversibility under precursor-level thermal competition.²⁶⁻²⁷

Motivated by this consensus, the study designs three controlled comparative groups to identify influence factors on lignin-driven pore-mouth regulation: (i) 1300 °C vs 1500 °C to evaluate carbon-intermediate structural-competition impact on pore-entrance sealing behaviour; (ii) pre-carbonised vs pristine YP-50F to quantify how lignin modifies entrance stability on the same scaffold; and (iii) pre-oxidised vs un-oxidised lignin to examine chemical-structure effects on pore-sealing efficiency and interfacial compatibility. XRD and Raman peak evolution are used to monitor microcrystalline evolution and defect trends, and half/full-cell tests are used to verify the ICE–plateau compatibility window. The objective throughout is to establish a structure–electrochemistry correlation that elucidates lignin’s sealing/modulation role at existing micropore entrances and to establish design principles for pore-mouth engineering to enhance SIB anode specific capacity.

2. Methodology

2.1 Materials

Organosolv lignin (Sigma-Aldrich) and activated carbon YP50F (Kuraray, Japan) were used as carbon precursors. Acetone ($\geq 99.5\%$) was used to dissolve lignin. Carboxymethyl cellulose (CMC) was employed as the binder, and aluminium foil (15 μm) was used as the current collector. For half-cell measurements, sodium metal (99.9%) served as the counter electrode, and the electrolyte consisted of 1 M NaPF₆ in EC/DEC (1:1 v/v).

2.2 Synthesis of Lignin–YP50F Hybrid Hard Carbon

This study evaluated three types of lignin–YP50F hybrids prepared under different conditions: (i)

pristine YP50F; (ii) pre-carbonised YP50F obtained by heating at 900 °C for 1 h under N₂ flow; (iii) hybrids prepared using pre-oxidised lignin.

Before mixing, YP50F powders were dry milled for 20 min to reduce particle agglomeration. For hybrid precursor preparation, the mass of YP50F was fixed at 0.5 g, and lignin was added to achieve YP50F: lignin mass ratios of 3:7, 4:6, 1:1, and 6:4. For each gram of total solid, 10 mL of acetone was added. Lignin was first dispersed in acetone under magnetic stirring for 20 min, after which YP50F was added and stirred for an additional 5 min, followed by ultrasonication for 30 min. The resulting suspension was dried overnight at 80 °C. The dried powders were then carbonised under N₂ flow (500 mL min⁻¹) using the following heating protocols:

900 °C for 1 h, followed by 1300 °C for 2 h, at a heating rate of 5 °C min⁻¹;

900 °C for 1 h, followed by 1500 °C for 2 h, at a heating rate of 5 °C min⁻¹.

For samples involving pre-oxidised lignin, lignin was pre-oxidised for 7 h at 250 °C prior to mixing.

After carbonisation, the materials were wet-milled with a small amount of isopropanol for 20 min, dried overnight, and subsequently dry-milled for 25 min. The resulting hybrid carbons were denoted as YP50F-L-T, where T represents the final carbonisation temperature (1300 or 1500 °C). The notation PO refers to pre-oxidised lignin, and PC refers to pre-carbonised YP50F. For example, (PC)3(PO)7-1500 represents a sample prepared by mixing pre-carbonised YP50F (3 parts by mass) with pre-oxidised lignin (7 parts by mass), followed by final carbonisation at 1500 °C.

2.3 Materials Characterisation.

The structural and textural properties of the carbon composites were characterised by nitrogen adsorption-desorption, X-ray diffraction (XRD), and Raman spectroscopy. Nitrogen adsorption-desorption measurements were performed at 77 K on a Micromeritics TriStar surface area and porosity analyser (Micromeritics Instrument Corp., USA) after degassing the samples at 220 °C for 16 h. The BET method was used to calculate the specific surface area from the nitrogen adsorption isotherms. XRD patterns were collected on a PANalytical X'Pert Pro MPD diffractometer (Malvern PANalytical, Netherlands) using Cu K α radiation ($\lambda = 1.5406$ Å) over a 2θ range of 10–80° to evaluate the microcrystalline structure of the carbon materials. Raman spectra were recorded on a WiTec Raman spectrometer (WiTec GmbH, Germany) equipped with a 532 nm excitation laser, and the ID/IG intensity ratio was used to assess the degree of structural disorder in the carbon framework.

For the analysis of XRD data, XRD patterns were baseline-corrected, and the (002) and (100) peaks were fitted using Gaussian peak models. The fitted peak positions and FWHM values were then used to

calculate d_{002} , L_c , L_a , and N according to the equations provided below.

$$d_{002} = \frac{\lambda}{2\sin\theta_{002}} \quad (\text{equ2.1})$$

$$L_c = \frac{K\lambda}{B_{002}\cos\theta_{002}} \quad (\text{equ2.2})$$

$$L_a = \frac{1.84\lambda}{B_{100}\cos\theta_{100}} \quad (\text{equ2.3})$$

$$N = \frac{L_c}{d_{002}} \quad (\text{equ2.3})$$

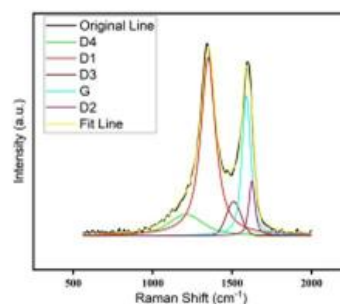


Figure 1. Five peak fitting for Raman Spectrum

Raman spectra (500–2000 cm⁻¹) were fitted using a five-peak deconvolution model based on the literature²⁸, including the D1, D2, D3, D4, and G bands (Figure 1). Peak fitting was performed using a Pseudo-Voigt (Gaussian-Lorentzian mixed) function with baseline correction to obtain the integrated peak areas for calculating the I_D/I_G ratios for further analysis.

2.4 Preparation of Working Electrodes

The hybrid hard carbon powders obtained during synthesis were subjected to a two-stage grinding process to ensure uniform particle size. In the first stage, the powders were wet milled with isopropanol for 20 min, followed by overnight drying in an oven. A second dry-grinding step was then performed for 25 min to further break down agglomerates, particularly due to the high hardness of lignin-derived carbon domains. After grinding, the powders were mixed with the binder carboxymethyl cellulose (CMC) at a mass ratio of 9:1 (active material: CMC). The slurry was homogenised and subsequently centrifuged for 3 min to remove entrapped air bubbles and improve coating uniformity. The mixture was then manually coated onto aluminium foil using a cylindrical rod as a doctor blade set to a thickness of 200 μm , to ensure a uniform coating layer. After drying, the electrodes were cut into square pieces with a dimension of 0.8 $\times$ 0.8 cm, corresponding to an active material loading of 2.4–3.0 mg cm⁻². The prepared electrodes were stored in a vacuum oven prior to coin-cell assembly to prevent moisture uptake.

2.5 Sodium Half-Cell Assembly and Electrochemical Measurements

Sodium half-cells were assembled inside an argon-filled glovebox with controlled atmosphere ($\text{H}_2\text{O} < 0.5$ ppm, $\text{O}_2 < 0.5$ ppm). CR2032-type coin cell cases were used for fabrication. Working electrodes prepared and dried as described in Section 2.4 were transferred into the glovebox and weighed before assembly. The coin cell stack was arranged in the following order: stainless steel case (bottom), 1 mm stainless steel spacer, GF/C glass fibre separator, the carbon working electrode (carbon layer facing the separator), and a commercial sodium metal disk (12 mm diameter) serving as both counter and reference electrode. Approximately 100 μL of electrolyte was dispensed onto the separator using a micropipette to ensure full wetting. The electrolyte consisted of 1 M NaPF_6 dissolved in EC: DEC (1:1 v/v) with 5 wt% FEC additives. A stainless-steel spring and the top case were positioned above the sodium electrode, and the coin cell was hermetically sealed using a hydraulic crimping machine. After sealing, the cells were immediately subjected to subsequent electrochemical testing. The electrochemical performance was evaluated using a theoretical capacity of 250 mAh g^{-1} to define the C-rates. The cells were cycled for five cycles each at 0.1, 0.2, 0.5, 1.0, and back to 0.1 C to assess rate performance at

different C-rates and the reversibility of the electrodes after fast charging and discharging.

3. Results and Discussion

The structural and electrochemical analysis began with non-oxidised lignin electrodes, and afterwards, a pre-oxidation step was introduced. Pre-oxidation promotes oxidative crosslinking of lignin, preventing thermoplastic flow and enabling controlled carbon deposition onto YP50. It stabilises the pore structure and converts many open micropores into closed, stable nano-cavities. This reduces defect density and SEI formation. Ultimately, it yields more reversible and stable Na-storage behaviour in lignin-rich electrodes. Hence, a 250°C-pre-oxidation step was introduced. Structural and electrochemical analyses were also performed on pre-oxidised lignin electrodes.

3.1. Characterisation of the Composite Structure.

The N_2 adsorption-desorption results provide clear insights into how precursor chemistry, lignin content, and the thermal history of YP50F collectively determine the efficiency of pore-mouth regulation in the composite carbons. Based on (Figure 2.a). All samples exhibit mixed type-I/IV isotherms, indicating that the materials retain the intrinsic

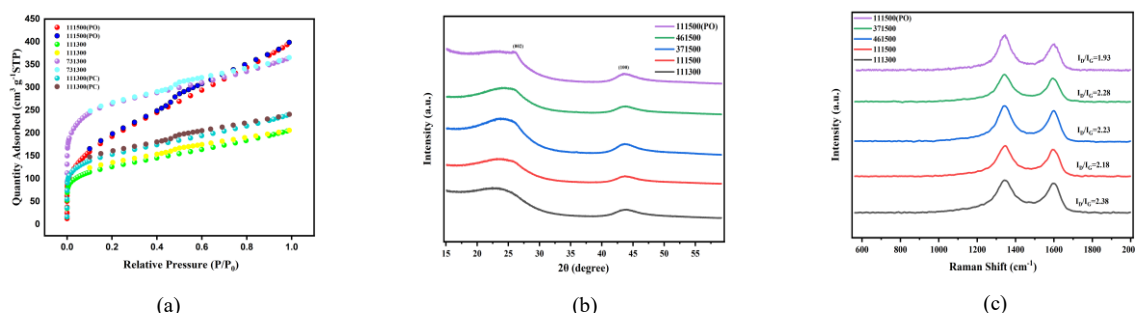


Figure 2. (a) N_2 adsorption isotherms of the YP50F–lignin composites. (b) XRD patterns showing turbostratic carbon structures. (c) Raman spectra with ID/IG ratios indicating defect levels.

Table 1. XRD structural parameters of the composite carbons.

Grouping	$d_{002}(\text{nm})$	$L_c(\text{nm})$	$L_a(\text{nm})$	N	FWHM (002)	FWHM (100)
111300	0.376924	1.095166	1.846451	2.905534	7.32802	4.34639
111500	0.370764	1.106725	2.205761	2.984982	7.25678	3.64104
371500	0.369135	1.15189	2.189705	3.12051	6.97364	3.66847
461500	0.366145	1.142968	2.267124	3.121625	7.0307	3.54452

micropore-filling characteristics of YP50F while also developing a certain degree of mesoporosity. The low-lignin sample 731300 shows the highest N_2 uptake and a large BET surface area of 956.78 $\text{m}^2 \text{g}^{-1}$, demonstrating that the lignin content is

insufficient to effectively constrict or seal the micropore entrances of YP50F. As lignin content increases or its chemical state changes, the accessible pore volume progressively decreases; however, the regulation effectiveness remains

strongly dependent on the intrinsic properties of the YP50F scaffold. For example, the pre-oxidized sample 111500(PO) still exhibits a relatively high BET surface area ($708.76 \text{ m}^2 \text{ g}^{-1}$), suggesting that although pre-oxidation enhances crosslinking within lignin, its carbonized products cannot form a continuous sealing layer, leaving a substantial fraction of micropores exposed.

A comparison between pre-carbonized and non-pre-carbonized YP50F reveals an interesting phenomenon: BET analysis shows that the hybrid carbon produced from pre-carbonised active carbon and lignin exhibits a higher specific surface area ($452.27 \text{ m}^2 \text{ g}^{-1}$). This indicates that pre-carbonisation of YP50F stabilises the micropore framework, but it may also reduce the subsequent pore-mouth filling and regulatory effect of lignin. A reasonable hypothesis is as follows: the 900°C pre-carbonization treatment prevents the YP50F pore walls from undergoing significant softening, deformation, and rearrangement during the second carbonization process, thereby making it difficult for lignin to enter, be compressed, or be sealed inside. At the same time, the ultra-micropores and internal defects introduced by pre-carbonisation lead to a stabilised composite carbonisation structure in which lignin cannot access the internal pore system, leaving these pores unsealed. In addition, pre-carbonisation may reduce the surface energy, leading to preferential deposit of lignin on the external surface. These synergistic effects enable PC-YP50F to generate more N_2 -accessible micropores after composite formation, resulting in a higher specific surface area. These hypotheses may be further verified in future studies through SEM to directly observe the structural features.

The differences in pore-mouth regulation are most clearly reflected by the variation in lignin content. We can compare 731300 and 111300. In 731300, adsorption and desorption curves overlap almost completely, and no observable hysteresis loop appears. This indicates that pore-mouth constriction is almost negligible, and the pore structure remains nearly fully open. Such a state corresponds to an unsealed pore condition, which results from insufficient lignin content that fails to effectively fill the micropores of YP50F. In contrast, 111300 shows a clear hysteresis loop. This indicates that partial pore narrowing has occurred, and pore sealing is relatively successful. Nitrogen can enter these sealed pores during adsorption but becomes restricted during desorption, demonstrating that its pore-mouth regulation effect is much stronger than that of 731300. Overall, we observe that the BET trend $731300 > 111500(\text{PO}) > 111300(\text{PC}) > 111300$ reflects the synergistic influence of lignin content, oxidation state, and YP50F pretreatment on

the accessible pore structure of the composite materials. The degree of hysteresis observed from nitrogen adsorption further illustrates the effectiveness of pore-mouth regulation.

The XRD patterns of the samples (Figure 2.b) exhibit the broad (002) and (100) reflections characteristic of turbostratic hard carbon. The fitted (002) peak positions yield interlayer spacings d_{002} in the range of $0.366\text{--}0.376 \text{ nm}$, calculated using the Bragg-based procedure described in the Methodology section. This spacing falls within the typical regime required for Na^+ insertion, confirming that all samples exhibit enlarged and irregular layer gaps suitable for sodium storage.

Similarly, the stacking height L_c , lateral crystallite size L_a , and microcrystalline layer number N were derived from the peak widths following the Scherrer-based approach detailed in Methodology (Table 1). Although these parameters inherently carry uncertainty due to the broad diffraction features of highly disordered carbons, the obtained values consistently indicate short stacking domains, limited in-plane ordering, and a highly turbostratic structure. Therefore, the XRD results mainly serve to confirm the structural framework and trend evolution of the materials, while finer distinctions in disorder and defect configurations are more reliably captured by Raman spectroscopy, as discussed in the next section.

The Raman Spectrum was integrated into a two-peak diagram (Figure 2.c) from the five-peak fitting method for easier analysis of the ID/IG ratio. For higher lignin ratio (from 1:1 to 3:7), the ID/IG ratio rises slightly from 2.18 to 2.28, indicating that the use of lignin introduces more defects to the hard carbon system. At higher carbonisation temperatures, the ID/IG ratio drops from 2.38 at 111300 to 2.18 at 111500, indicating that higher carbonisation temperature leads to a higher degree of order. Also, the introduction of the pre-oxidation step leads to a drop in ID/IG, showing it decreases defect sites brought by lignin and yields a more stable carbon morphology, consistent with the expectation for a pre-oxidation step.

3.2. Electrochemical Properties of Composite Materials.

3.2.1 Electrochemistry Analysis of non-preoxidised samples

Because 1500°C provides a better pore-closing effect than 1300°C , only electrodes with a carbonization temperature of 1500°C were prepared. The electrochemical effects of different lignin ratios and lignin peroxidation were analysed as well.

Based on the dQ/dV curves in Figure 4, the sodium-storage behaviours of the three hard carbon

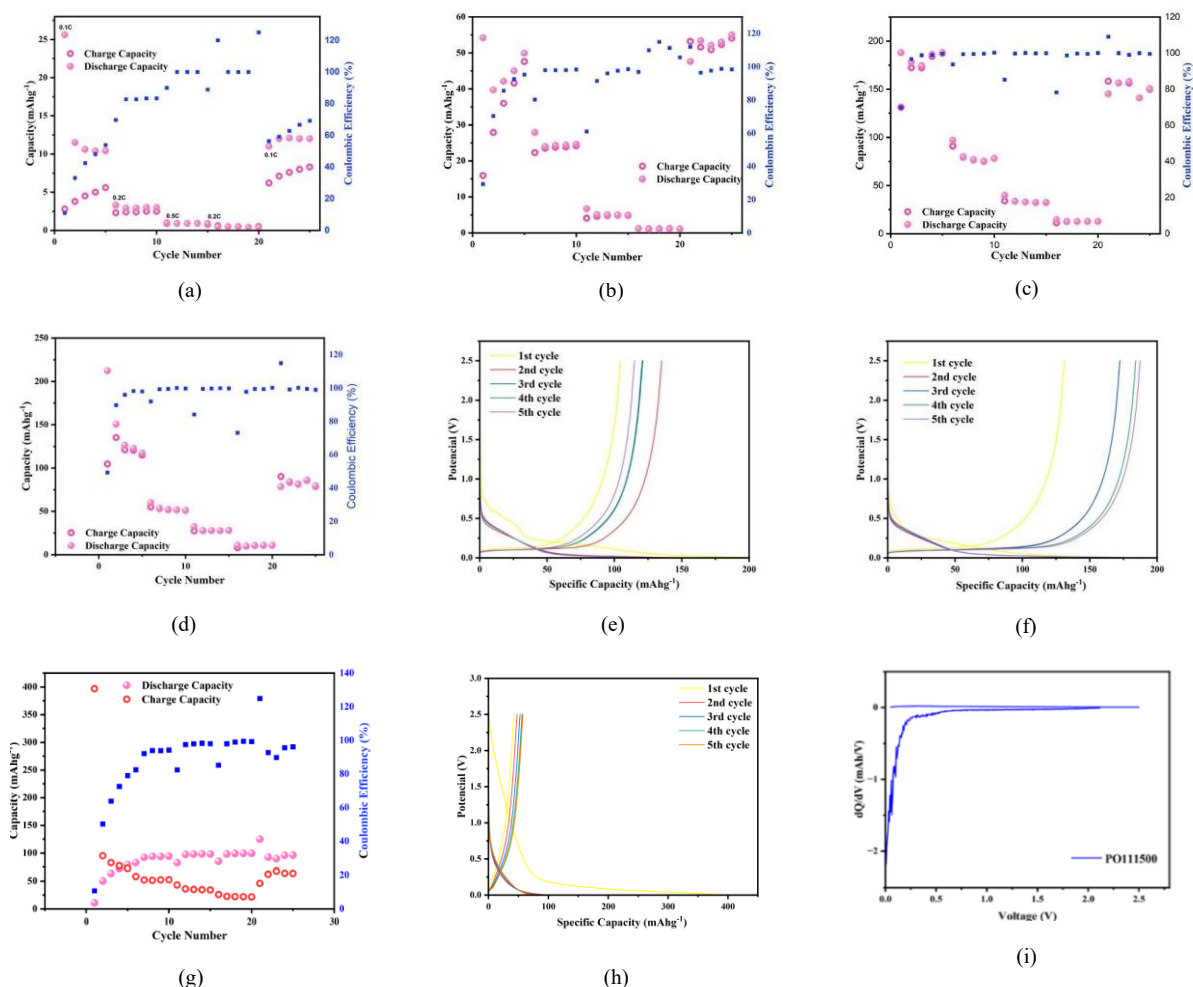


Figure 3. Cycling performance and voltage profiles of composite hard carbons. (a) Sample 371300 (3YP50F-7lignin, 1300 °C). (b) 371500. (c) 461500. (d) 111500 cycling performances. (e) Charge-discharge curves of 111500. (f) Charge-discharge curves of 461500. (g) PO111500 cycling performance. (h) Charge-discharge curves of PO111500. (i) dQ/dV Curve of PO111500.

samples can be divided into the low-voltage region (0-0.1V) and the high-voltage region (0.1V and above). The low-voltage region serves as the pore-filling stage for quasimetallic Na. The high-voltage region results from the filling of defects and the interlayer of hard carbon. The introduction of lignin mainly controls pore size, and pore filling provides the primary source of capacity. Hence, in this paper, we will analyse only the low-voltage region.

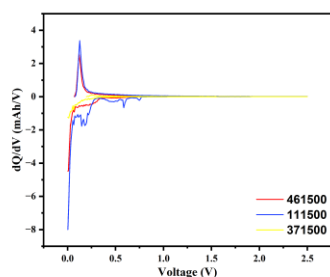


Figure 4. dQ/dV Curve of non-preoxidised samples

In the low-voltage region, all samples exhibit a pronounced cathodic peak associated with pore filling in closed/ultra-microporous domains and SEI formation. The peak intensity follows the order 111500 (Figure 3.d) > 461500 (Figure 3.c) > 371500 (Figure 3.b). The 111500 sample shows the sharpest and deepest peak, indicating the largest population of open pores and the highest surface area, which leads to stronger initial Na⁺ and electrolytes filling and significant interfacial reactions. These might lead to a lower ICE and a larger irreversible capacity, as well as a drop in overall extractable capacity. The 461500 sample displays a moderate peak, suggesting an intermediate pore structure, while the 371500 sample exhibits the weakest low-voltage response, implying fewer open pores and hence a milder interfacial reaction—beneficial for higher initial Coulombic efficiency.

To further assess the pore-closing effect in an actual battery design, a five-cycle test with different charge and discharge rates (0.1C, 0.2C, 0.5C, 1C,

and back to 0.1C) was performed to evaluate the cycling stability and reversibility of each hard carbon anode (Figure 3. e&f). As expected, 4:6 provides a higher CE of nearly 80% compared to 1:1 at 60%, and recovers most of the capacity after high-C-rate cycles due to a better pore-closing effect and reduced SEI side reactions. This could be further demonstrated by the higher accessible capacity of 185 mAh/g in 461500 compared to 130 mAh/g in 111500. These show that pore size is controlled by additional lignin, leading to a higher ICE, reduced SEI side reactions, and better reversible capacity. However, in contrast to the stereotype that more lignin provides higher ICE, better capacity, and greater reversibility, the overall accessible capacities of 371500 and 371300 are very low (<50 mAh/g) even at low C-rates. One possible assumption is that the great amount of lignin insulates the pores too much, blocking the pore-filling pathways of sodium ions. In this case, almost no pore filling happens. This is demonstrated by the lack of capacity change in the dQ/dV diagram, indicating that only a few sodium ions can be filled in pores and limited capacity can be extracted from the 3:7 anodes.

Overall, increasing the YP50 content (from 3:7 to 1:1) enhances both the pore-filling contribution at low voltages and the defect-associated reactions at higher voltages, providing more Na-storage sites but also intensifying side reactions and polarisation. Hence, a balance between the number of pores and pore size must be achieved. It is observed that increasing the lignin ratio to 60% achieves the best pore-closing, preventing most side SEI reactions and ensuring the filling of Sodium-ion. In contrast, further increasing the lignin content to 70% should exhibit the most stable and reversible electrochemical behaviour, as it has a lower defect density, reduced open-pore volume, and lower specific surface area, in theory. However, excess lignin blocks filling pathways for sodium, resulting in a low capacity even at low C-rates. Consequently, the 4:6 ratio is determined to be the best for pore closure.

3.2.2 Electrochemistry Analysis of preoxidised samples

A pre-oxidised lignin: YP50 (1:1) sample was prepared, and its electrochemical properties were measured with the same methods. The dQ/dV profile (Figure 3.i) exhibits no discernible low-voltage peaks—similarly to that of the unoxidised 3:7 sample. In comparison with the non-preoxidised carbons (1:1 and 4:6), the pre-oxidised material delivers a markedly lower first-cycle capacity, shows almost no capacity increase in subsequent cycles over the first round, and maintains excellent reversibility after high C-rate (up to 1 C) operation (Figure 3. gh). These behaviours are consistent with a significantly reduced population of accessible micropores. Pre-oxidation removes many

micropores and defect sites while generating a larger fraction of closed pores, of which only a limited proportion can be accessed by Na⁺. This greatly decreases the overall reversible capacity. Nevertheless, the more ordered and uniform carbon framework ensures that nearly all Na⁺ pore filling occurs through stable closed-pore domains, effectively suppressing parasitic reactions and ensuring a stable SEI formation. As a result, the initial capacity approaches the maximum achievable value. The strong reversibility further demonstrates that the closed-pore-dominated structure successfully limits side reactions even under high-rate cycling. Overall, these results indicate that pre-oxidation is an effective strategy for tuning pore structure. Future optimisation may focus on designing closed-pore architectures that remain accessible to Na⁺ while still mitigating interfacial degradation. Given that peroxidation provides much better pore closure with the same amount of lignin, a possible solution is to reduce the amount of lignin added.

4. Conclusion and Outlook

In structural analysis, 1500 °C treatment results in stronger pore-mouth constriction and a reduced interlayer spacing compared with 1300 °C. Non-precarbonised samples also show greater entrance contraction than pre-carbonised ones. Based on these observations, the optimal lignin ratio for non-preoxidised samples was determined under 1500 °C carbonisation. Low lignin content does not sufficiently constrict pore mouths and leads to parasitic reactions, while excessively high lignin content over-contracts the entrances and generates a large fraction of Na⁺-inaccessible (blocked) pores. An intermediate lignin ratio achieves a balance between Na⁺ accessibility and parasitic-reaction suppression, enabling efficient pore filling and improved ICE. Some degree of side reactions remain, indicating that pore-mouth regulation can be further refined.

Preoxidation further strengthens pore-mouth constriction and enhances structural stability. Compared with the unpreoxidised counterpart, the preoxidised lignin composite shows more pronounced entrance contraction, reduced defect density, and improved cycling reversibility. However, the maximum accessible capacity is limited due to excessive entrance blocking, suggesting that over-constriction suppresses Na⁺ diffusion into internal closed-pore domains.

For future improvement, high-resolution TEM may provide indirect evidence for nanoscale structural evolution associated with pore-mouth constriction and closed-pore formation. Solid-state NMR could help clarify how lignin functional-group chemistry influences cross-linking and pore-sealing behaviour, offering deeper insight into lignin preoxidation mechanisms. In addition, tuning the

amount of preoxidised lignin may provide a pathway to simultaneously suppress parasitic reactions and maintain sufficient Na⁺ accessibility for pore filling.

5. Acknowledgment

We would like to thank Prof. Magda Titirici and Mr. Joe (Yijun) Yang for their continuous guidance throughout this project. We also thank Mr. Yiliang Shan and Mr. Yichen Huang for providing training on materials characterization and battery assembly, and Dr. Kalpana Balaskandan for her instruction and assistance with glovebox operation.

6. References

- [1] International Energy Agency (2023). *Trends in batteries – Global EV Outlook 2023 – Analysis*. [online] IEA. Available at: <https://www.iea.org/reports/global-ev-outlook-2023/trends-in-batteries>.
- [2] Nations, U. (2021). *Frontier Technology Issues: Lithium-ion batteries: a pillar for a fossil fuel-free economy? | Economic Analysis and Policy Division*. [online] United Nations. Available at: <https://policy.desa.un.org/publications/frontier-technology-issues-lithium-ion-batteries-a-pillar-for-a-fossil-fuel-free>.
- [3] Waseem, M., Lakshmi, G.S., Ahmad, M. and Suhaib, M. (2024). Energy storage technology and its impact in electric vehicle: Current progress and future outlook. *Next Energy*, [online] 6, pp.100202–100202. doi:<https://doi.org/10.1016/j.nxener.2024.100202>.
- [4] Navid Nasajpour-Esfahani, Hamid Garmestani, Mohsen Bagheritabar, Jasim, D.J., D. Toghraie, Shohreh Dadkhah and Hooman Firoozeh (2024). Comprehensive review of lithium-ion battery materials and development challenges. *Renewable and Sustainable Energy Reviews*, 203, pp.114783–114783. doi:<https://doi.org/10.1016/j.rser.2024.114783>.
- [5] Nayak, P.K., Yang, L., Brehm, W. and Adelhelm, P. (2018). From Lithium-Ion to Sodium-Ion Batteries: Advantages, Challenges, and Surprises. *Angewandte Chemie (International ed. in English)*, [online] 57(1), pp.102–120. doi:<https://doi.org/10.1002/anie.201703772>.
- [6] Larcher, D. and Tarascon, J-M. (2014). Towards greener and more sustainable batteries for electrical energy storage. *Nature Chemistry*, [online] 7(1), pp.19–29. doi:<https://doi.org/10.1038/nchem.2085>.
- [7] Li, X., Ding, C., Liang, Q., Hu, J., Xu, L., Li, Y., Liu, Y. and Gao, Y. (2024). Progress in hard carbons for sodium-ion batteries: Microstructure, sodium storage mechanism and initial Coulombic efficiency. *Journal of Energy Storage*, [online] 98, p.112986. doi:<https://doi.org/10.1016/j.est.2024.112986>.
- [8] Dou, X., Hasa, I., Saurel, D., Vaalma, C., Wu, L., Buchholz, D., Bresser, D., Komaba, S. and Passerini, S. (2019). Hard carbons for sodium-ion batteries: Structure, analysis, sustainability, and electrochemistry. *Materials Today*, 23, pp.87–104. doi:<https://doi.org/10.1016/j.mattod.2018.12.040>.
- [9] Chu, Y., Zhang, J., Zhang, Y., Li, Q., Jia, Y., Dong, X., Xiao, J., Tao, Y. and Yang, Q. (2023). Reconfiguring Hard Carbons with Emerging Sodium-ion Batteries: a Perspective. *Advanced Materials*, p.2212186. doi:<https://doi.org/10.1002/adma.202212186>.
- [10] Qiu, C., Li, A., Qiu, D., Wu, Y., Jiang, Z., Zhang, J., Xiao, J., Yuan, R., Jiang, Z., Liu, X., Chen, X. and Song, H. (2024). One-Step Construction of Closed Pores Enabling High Plateau Capacity Hard Carbon Anodes for Sodium-Ion Batteries: Closed-Pore Formation and Energy Storage Mechanisms. *ACS Nano*, 18(18), pp.11941–11954. doi:<https://doi.org/10.1021/acsnano.4c02046>.
- [11] Huang, Z., Qiu, X., Wang, C., Jian, W., Zhong, L., Zhu, J., Zu, X. and Zhang, W. (2023). Revealing the effect of hard carbon structure on the sodium storage behavior by using a model hard carbon precursor. *Journal of energy storage*, 72, pp.108406–108406. doi:<https://doi.org/10.1016/j.est.2023.108406>.
- [12] Zhang, T., Li, C., Wang, F., Noori, A., Mousavi, M.F., Xia, X. and Zhang, Y. (2022). Recent Advances in Carbon Anodes for Sodium-Ion Batteries. *The Chemical Record*, 22(10). doi:<https://doi.org/10.1002/tcr.202200083>.
- [13] Chen, X., Liu, C., Fang, Y., Ai, X., Zhong, F., Yang, H. and Cao, Y. (2022). Understanding of the sodium storage mechanism in hard carbon anodes. *Carbon Energy*. doi:<https://doi.org/10.1002/cey2.196>.
- [14] Sun, N., Qiu, J. and Xu, B. (2022). Understanding of Sodium Storage Mechanism in Hard Carbons: Ongoing Development under Debate. *Advanced Energy Materials*, 12(27), pp.2200715–2200715. doi:<https://doi.org/10.1002/aenm.202200715>.
- [15] Li, Y., Lu, Y., Zhao, C., Hu, Y.-S., Titirici, M.-M., Li, H., Huang, X. and Chen, L. (2017). Recent advances of electrode materials for low-cost sodium-ion batteries towards practical application for grid energy storage. *Energy Storage Materials*, 7, pp.130–151. doi:<https://doi.org/10.1016/j.ensm.2017.01.002>.
- [16] Zhao, D., Chen, C., Zhang, Q., Chen, W., Liu, S., Wang, Q., Liu, Y., Li, J. and Yu, H. (2017). High Performance, Flexible, Solid-State Supercapacitors Based on a Renewable and Biodegradable Mesoporous Cellulose Membrane. *Advanced Energy Materials*, 7(18), p.1700739. doi:<https://doi.org/10.1002/aenm.201700739>.
- [17] Wang, T., Huang, C., Zhou, X. and Zhan, Y. (2025). A novel type IV deep eutectic solvent for pretreating wheat straw to co-produce fermentable sugars and hydrophobic micro-nano lignin particle. *Bioresource Technology*, pp.133739–

133739.
doi:<https://doi.org/10.1016/j.biortech.2025.133739>.
- [18] Dou, X., Hasa, I., Maral Hekmatfar, Diemant, T., Rolf Jürgen Behm, Buchholz, D. and Khademhosseini, A. (2017). Pectin, Hemicellulose, or Lignin? Impact of the Biowaste Source on the Performance of Hard Carbons for Sodium-Ion Batteries. *Chemsuschem*, 10(12), pp.2668–2676. doi:<https://doi.org/10.1002/cssc.201700628>.
- [19] Kim, J., Han, S.-D., Koo, B., Lee, S.-H. and Yang, J. (2023). Structure Dependent Electrochemical Behaviors of Hard Carbon Anode Materials Derived from Natural Polymer for Next-Generation Sodium Ion Battery. *Polymers*, 15(22), p.4373. doi:<https://doi.org/10.3390/polym15224373>.
- [20] Wang, A., Zhang, G., Li, M., Sun, Y., Tang, Y., Sun, K., Lee, J.-M., Fu, G. and Jiang, J. (2025). Lignin derived hard carbon for sodium ion batteries: Recent advances and future perspectives. *Progress in Materials Science*, [online] 152, p.101452. doi:<https://doi.org/10.1016/j.pmatsci.2025.101452>.
- [21] Du, Y.-F., Sun, G.-H., Li, Y., Cheng, J.-Y., Chen, J.-P., Song, G., Kong, Q.-Q., Xie, L.-J. and Chen, C.-M. (2021). Pre-oxidation of lignin precursors for hard carbon anode with boosted lithium-ion storage capacity. *Carbon*, 178, pp.243–255. doi:<https://doi.org/10.1016/j.carbon.2021.03.016>.
- [22] Chen, B., Zhong, L., Lu, M., Jian, W., Sun, S., Meng, Q., Wang, T., Zhang, W. and Qiu, X. (2024). Lignin molecular sieving engineering enables high-plateau-capacity hard carbon anodes for sodium-ion batteries. *Green Chemistry*, 26(13), pp.7919–7930. doi:<https://doi.org/10.1039/d4gc02019g>.
- [23] Wang, H., Sun, F., Qu, Z., Wang, K., Wang, L., Pi, X., Gao, J. and Zhao, G. (2019). Oxygen Functional Group Modification of Cellulose-Derived Hard Carbon for Enhanced Sodium Ion Storage. *ACS Sustainable Chemistry & Engineering*, 7(22), pp.18554–18565. doi:<https://doi.org/10.1021/acssuschemeng.9b04676>.
- [24] Narasimharao Kitchamsetti and Sungwook Mhin (2025). Advances and Prospects of Lignin-Derived Hard Carbons for Next-Generation Sodium-Ion Batteries. *Polymers*, 17(20), pp.2801–2801. doi:<https://doi.org/10.3390/polym17202801>.
- [25] Zhao, L., Sun, Z.-F., Zhang, C.-C., Nan, J., Ren, N.-Q., Lee, D.-J. and Chen, C. (2022). Advances in pretreatment of lignocellulosic biomass for bioenergy production: Challenges and perspectives. *Bioresource Technology*, 343, p.126123. doi:<https://doi.org/10.1016/j.biortech.2021.126123>.
- [26] Tian, Y.-R., Yi, Z.-L., Su, F.-Y., Xie, L.-J., Zhang, X.-F., Li, X.-F., Cheng, J.-Y., Chen, J.-P. and Chen, C.-M. (2024). Regulating the Pore Structure of Activated Carbon by Pitch for High-Performance Sodium-Ion Storage. *ACS Applied Materials & Interfaces*, 16(14), pp.17553–17562. doi:<https://doi.org/10.1021/acsami.4c00301>.
- [27] Zhang, Y., Zhang, S.-W., Chu, Y., Zhang, J., Xue, H., Jia, Y., Cao, T., Qiu, D., Zou, X., Wang, D.-W., Tao, Y., Zhong, G., Peng, Z., Kang, F., Lv, W. and Yang, Q.-H. (2025). Redefining closed pores in carbons by solvation structures for enhanced sodium storage. *Nature Communications*, 16(1). doi:<https://doi.org/10.1038/s41467-025-59022-8>.
- [28] Sadezky, A., Muckenhuber, H., Grothe, H., Niessner, R. and Pöschl, U. (2005). Raman microspectroscopy of soot and related carbonaceous materials: Spectral analysis and structural information. *Carbon*, 43(8), pp.1731–1742. doi:<https://doi.org/10.1016/j.carbon.2005.02.018>.

Targeted nanomedicine for effective thrombolysis

Xiquan Gong and Xuan Gan

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Streptokinase (SK) is widely used for thrombolysis but primarily targets fibrin degradation, which is often insufficient for platelet-rich clots where strong platelet–platelet interactions dominate clot stability. In this project, red blood cell–derived vesicles (RBCVs) were explored as biomimetic nanocarriers to co-deliver SK with ascorbic acid (AA), introduced as an adjuvant to weaken platelet interactions by exerting a redox stress.

RBCVs were prepared from human red blood cells and characterised to elaborate their structural properties. To quantify encapsulated AA, an AA-DCPIP assay was optimised, and the calibration curve was shown to remain linear and reliable in the presence of SK and residual haemoglobin found in RBCVs. Using the calibration curve, the encapsulation efficiency of AA within RBCVs was found for different protocols, including incubation time and surface modification. The thrombolytic potential of the nano formulations was tested using a halo-clot assay. AA-loaded RBCVs combined with SK achieved faster and more extensive clot degradation than free SK despite donor variability. These findings suggest that AA can complement conventional fibrin-targeting thrombolytics by providing dual-mechanism thrombolysis, and that RBCVs provide a promising platform for dual-mechanism thrombolytic delivery.

Key words: Encapsulation Efficiency, Thrombolysis, Nanoparticles, Ascorbic Acid, Red Blood Cell Vesicles

1 Introduction

Blood continuously circulates throughout the human body, delivering oxygen and essential nutrients to tissues while removing metabolic waste to sustain cellular function. However, when vascular injury or dysregulated coagulation occurs, blood clots (thrombi) can form within blood vessels, obstructing normal blood flow. This may lead to severe clinical conditions such as ischemic stroke, myocardial infarction, or deep vein thrombosis. These diseases cause serious consequences for society and are estimated to account for 1 in 4 deaths worldwide (Wendelboe & Raskob, 2016).

Thrombi are heterogeneous composites of fibrin, platelets, red blood cells, leukocytes and NETs, with fibrin fibres and platelet–platelet aggregates forming the principal structural backbone that governs clot stability and responsiveness to thrombolysis (Alkharithi, Duval, Yu, Macrae, & Ariëns, 2021). Traditional treatments for thrombosis typically rely on systemic administration of thrombolytic agents, such as tissue plasminogen activator (tPA) and streptokinase (SK). While clinically effective, these drugs are limited by short circulation time, low site selectivity, and a high risk of systemic bleeding, particularly due to off-target fibrinolysis. Given these limitations, there is an increasing need for strategies that enable targeted, timed, and quantitative thrombolytic release while minimising systemic exposure. To enhance the therapeutic precision and achieve these expectations, recent studies have explored nanoparticle-based delivery system to increase local drug concentration at the thrombosis site (Liu C.-H. , Rethi, Weng, Nguyen, & Chuang, 2024).

In addition, conventional thrombolytic agents predominantly target fibrin, so clots that are

particularly platelet-rich tend to be relatively resistant to lysis (Tomkins, et al., 2015).

2 Background

2.1 The Formation of Thrombosis

Under normal physiological conditions, haemostasis tightly regulates the balance between pro-coagulant and anticoagulant forces to preserve vascular integrity. When blood vessels are injured, platelets act as first responders, which adhere to the exposed subendothelial surface and, upon activation, release granules to recruit additional platelets (Davi & Patrono, 2007). Concurrently, the coagulation cascade is activated, culminating in the generation of thrombin, a key enzyme that converts fibrinogen into fibrin monomers, which polymerise and are stabilised by factor XIIIa to form a fibrin mesh that traps RBCs and platelets to seal the wound. Thrombin converts soluble fibrinogen into insoluble fibrin monomers, which then self-assemble into polymerized fibrin fibres. These fibres are further stabilised by cross-linking with factor XIIIa, forming a robust meshwork that traps RBCs and platelets to seal the wound (Désiré, 2001), which is briefly explained in the Figure1.1 below.

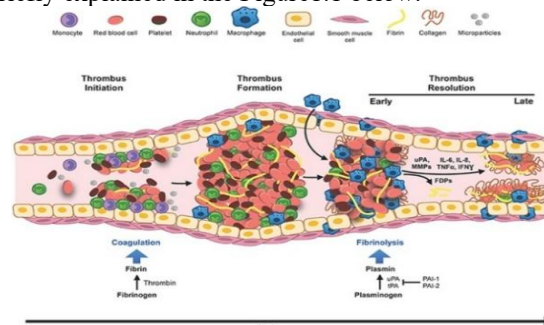


Figure1. 1: Innate immune cells in DVT (Subhradip Mukhopadhyay, 2019)

However, when this finely regulated mechanism becomes dysregulated, which is due to endothelial dysfunction, inflammation, slow blood flow, or hypercoagulable states, thrombus formation can occur in the absence of overt vascular injury (Cines, 1998). Such pathological thrombosis disrupts blood flow and can precipitate life-threatening events (Désiré, 2001), highlighting the continuing need for safer and more precise thrombolytic strategies.

Thrombosis is a central pathological process underlying numerous cardiovascular and cerebrovascular diseases including stroke and myocardial infarction.

2.2 Delivery System

Among the various biomimetic nanocarriers under investigation, red blood cell-derived vesicles (RBCVs) are particularly attractive for thrombolytic delivery because they preserve key biological features of native erythrocyte membranes while providing a versatile nanoscale carrier platform. RBC membranes enable immune evasion by presenting self-recognition proteins such as CD47 and multiple complement-regulatory factors, which suppress macrophage uptake and inhibit complement activation, thereby prolonging systemic circulation (Xia, Zhang, Li, Hou, & Feng, 2019). RBC-derived vesicles and membrane-coated nanocarriers also inherit the inherent biocompatibility and long bloodstream lifespan of native erythrocytes approximately 120 days, and ability to encapsulate or surface-load therapeutic cargos (Nguyen, Jayasinghe, Le, Peng, & Le, 2023).

2.3 Nanoparticle-Based Delivery Strategies for Targeted Thrombolysis

Currently used thrombolysis agents such as SK and tPA can effectively dissolve fibrin clots but often cause systematic bleeding due to poor targeting specificity. To enhance the therapeutic precision and overcome these issues, recent studies have explored nanoparticle-based delivery systems to increase local drug concentration at the thrombosis site (Liu C.-H. , Rethi, Weng, Nguyen, & Chuang, 2024).

tPA is a fibrin-targeting serine protease that attaches to the clot surface, activates fibrin-bound plasminogen, and generates plasmin to cleave the fibrin network (Jilani & Siddiqui, 2023). Streptokinase is a protein derived from the *Streptococcus* bacteria and is a potent plasminogen activator that converts the zymogen plasminogen into the fibrin clot-dissolving serine proteinase plasmin (Collen D, 1995). Additionally, the substantially lower cost of SK makes it a more practical choice when efficacy is comparable (White, 1989). Nevertheless, its lack of thrombus targeting and immunogenicity during circulation remain major challenges (T V Hrynenko 1, 2002). Consequently, there is a compelling need to develop improved, targeted SK delivery strategies that can

protect SK from degradation, enhance its localisation to the thrombus and reduce systemic exposure. Among the nanoparticle-based approaches, RBCVs have emerged as a promising biomimetic carrier capable of prolonging circulation time and evading immune clearance (Liu C.-H. , Rethi, Weng, Nguyen, & Chuang, 2024). Therefore, in this project, RBCVs are investigated as nanocarriers for SK with co-loading of antioxidants such as ascorbic acid (AA), with the aim of reducing oxidative platelet activation, enhancing targeted thrombolytic activity and potentially improve oxidative stability (Agnieszka & Skrzydlewska , 2022).

2.4 Ascorbic acid enhances thrombolysis as an adjuvant to SK

Oxidative stress is now recognised as a key regulator of platelet-dependent thrombus formation, comprising both reactive oxygen species (ROS) generation elicited by platelets during activation and responding to changes in redox balance (Freedman, 2008). Antioxidants like AA can modulate this pathway and have been reported to reduce platelet activation and aggregation (Tymł, 2017) (Liu, Pei, Hu, Gu, & Cui, 2020).

Antioxidants such as AA are therefore of interest not only as general cytoprotective agents but also as modulators of platelet–platelet interactions (Arauna, et al., 2024). Circulating AA has been shown to enhance nitric oxide-mediated anti-aggregatory signalling and to prevent platelet activation in vitro, thereby helping to keep platelets in a quiescent state and interrupting platelet–platelet aggregation (Raghavan, Sharma, & Dikshit, 2003) (Ellis, et al., 2001). Clinical and experimental studies further suggest that vitamin C can scavenge reactive oxygen species (ROS), reduce platelet activation markers and inhibit aggregation under pro-thrombotic or hypoxia/reoxygenation conditions (Liu, Pei, Hu, Gu, & Cui, 2020). These findings support the rationale for incorporating AA into thrombolytic delivery systems, with the aim of combining clot dissolution with local antioxidant and anti-platelet effects (Li, et al., 2022).

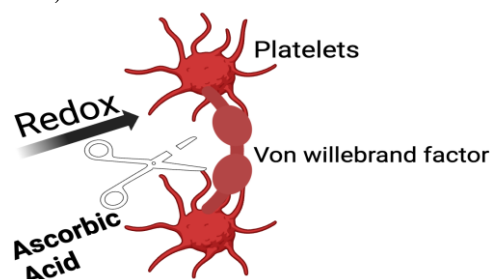


Figure 2.4.1 Working mechanism of Ascorbic acid

3 Methodology

3.1 Preparation of Reagents

Phosphate-buffered saline (PBS) (pH 7.4, Gibco) was used as the isotonic buffer throughout all

experiments, which could provide physiological ionic strength and buffering capacity.

Earlier studies in the Chen group recognised 12.5 $\mu\text{g/mL}$ as the optimal adjuvant concentration for AA. An AA stock (12.5mg/mL) solution was prepared and stored at 4 °C in the fridge. A 2,6-dichlorophenolindophenol (DCPIP) stock (19.03mg/mL) was prepared and kept in the dark by wrapping it in tin foil to prevent light induced degradation. To ensure experimental reliability, all solutions were replaced with freshly prepared stocks on a weekly basis. Prior to use, solutions were gently mixed to ensure uniform concentration and checked visually to confirm the absence of turbidity, colour change or sedimentation.

Serial dilutions of AA were freshly prepared immediately before the titration process, while DCPIP was diluted in the reservoir just prior to use under continuous light protection. All solutions were handled using calibrated micropipettes to ensure accuracy and reproducibility.

3.2 Quantification of ascorbic acid

To accurately determine the amount of AA encapsulated within RBCVs, a colourimetric quantification method based on the AA–DCPIP redox reaction was established. AA and DCPIP react in a 1:1 molar ratio. The principle as shown in the following Figure 3.2.1.

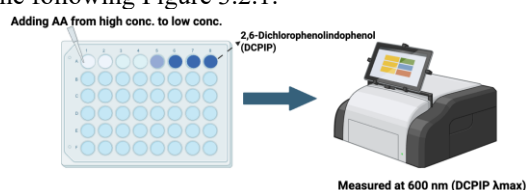


Figure 3.2.1 Schematic Illustration of DCPIP Reduction by Ascorbic Acid (Color Change from Blue to Colorless)

Since the working concentration of AA was in the microgram range, the optimal DCPIP titrant concentration had to be found which ensured the reaction progress to be captured in a reasonable time frame. The final titrant concentration of DCPIP was standardised to 190.3 $\mu\text{g/mL}$.

For each calibration point, AA and DCPIP were mixed in a 96-well microplate. Absorbance was measured at the maximum wavelength of DCPIP at 600 nm using the plate reader. DCPIP only and PBS only controls were included to correct for baseline absorbance.

The resulting calibration curve was used to define the linear analytical range of the AA–DCPIP reaction and was used for all subsequent analyses of AA recovered from purified RBCVs following loading procedures.

3.3 Preparation of Ghost Cells (GCs)

Ghost cell (GC) production was performed following the protocol reported by Henry (Henry, 2022).

GCs were prepared from human whole blood through sequential washing and hypotonic lysis

steps. All procedures involved directly contacting blood were performed inside a biosafety cabinet, and surfaces were disinfected with 70% ethanol before taken into the hood.

Whole blood was centrifuged at 150 x g for 10 minutes at 4 °C to separate plasma and cellular components from RBCs without activating platelets. After centrifugation, the plasma layer was carefully removed using a pipette. RBCs were washed 3 times in PBS to eliminate residual proteins. RBCs were suspended in PBS by inverting the tube with cells 10 times and centrifuging under the same conditions as mentioned above. Washed RBCs were then treated with EDTA–PBS (90% v/v) to induce hypotonic release of haemoglobin, where EDTA also acted as a calcium chelator to prevent coagulation. The mixtures were inverted gently to avoid foaming. Soon after, 10×PBS corresponding (10% v/v) was added to return the cells to an isotonic environment and the suspension was centrifuged at 12,100 rpm for 7 minutes at 4 °C. This wash–centrifuge cycle was repeated seven times to obtain purified GCs. The resulting pale pink pellet was washed once more with PBS and stored at 4 °C for subsequent use. Purified GCs were stored no longer than 2 weeks to ensure optimal membrane integrity.

The total protein content of the GC was quantified using the Bicinchoninic Acid (BCA) assay (ThermoFisher Scientific). The BCA working reagent was freshly prepared by mixing reagents A and B in a 50:1 ratio according to the instructions. The samples were incubated with the working reagent for 30 minutes at 37 °C to allow for color development through the reduction of Cu^{2+} to Cu^{+} by protein molecules, followed by the formation of a purple complex between Cu^{+} and BCA. After incubation, the absorbance of each well was measured at 562 nm. The average absorbance values were used to determine the protein concentration of the GC samples based on a standard calibration curve. After calculations, GC solution was suspended to 2200 $\mu\text{g/mL}$. The overview of BCA assay is shown in the following Figure 3.3.1.

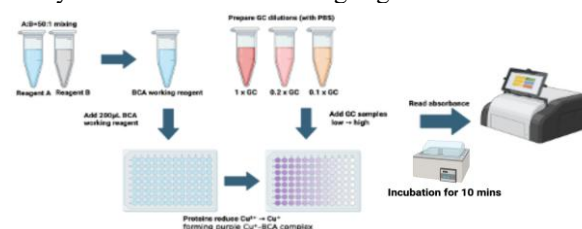


Figure 3.3.1 Workflow of BCA assay for GCs' quantification

3.4 Incubation and Membrane Modification

Incubation allowed components to be encapsulated within the ghost cell membranes through passive diffusion. RBCV formulations were made with 50% v/v GC (2200 $\mu\text{g/mL}$ stock, resulting in 1100 $\mu\text{g/mL}$ in the final mixture), along with drug of interest. For the surface modified RBCV formulation, 84.5 $\mu\text{L/mL}$ of cyclic Arginine–Glycine–Aspartic acid

peptide (cRGD) dissolved in dimethyl sulfoxide (DMSO) (BioPharma PEG) was added. cRGD was incorporated to provide targeting capability, as it specifically binds to integrin receptors (e.g., $\alpha\beta_3$) that are highly expressed on activated platelets. This enhances localization of RBCVs to thrombus sites. For the non-targeted formulation, the same volumes were used except that PBS was added in place of the cRGD solution.

All the mixture, ready to be incubated, was put into a flask with a small stirrer and was covered by a lid to prevent evaporation and contamination. Flasks were placed in a 37 °C water bath under gentle agitation. The total time taken for the incubation process is 1 hour, allowing sufficient interaction between the GCs and encapsulated components. During incubation, all liquids inside flasks were fully submerged, providing stable thermal contact and minimising local temperature fluctuations. The mixtures were collected for further sonication

3.5 Nanoparticles Formation

To achieve uniform nanoscale dispersion, the prepared samples were subjected to probe sonication. Each vial was placed in an ice-filled beaker to prevent overheating. Sonication was performed with a 1 second on / 1 second off cycle at 50% amplitude and 1800 Hz for a total of 20 minutes, giving an effective sonication time of 10 minutes. After treatment, samples were cooled and stored at 4 °C for later purification.

Previous studies highlight that AA is not significantly impacted by ultrasonic exposure (Aguilar, Garvín, Ibarz, & Augusto, 2017). Therefore, probe sonication was selected as a suitable method for nanoparticle formation in this study. The overview of nanoparticles preparation is illustrated in Figure 3.5.1 below.

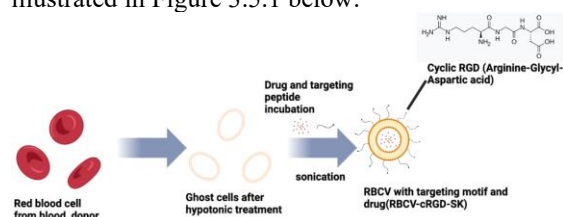


Figure 3.5.1 Preparation of cRGD-functionalised, SK-loaded RBCV

3.6 Characterisation of Nanoparticles

The size and polydispersity index (PDI) of the nanoparticles were determined using dynamic light scattering (DLS). Malvern Zetasizer was used for this measurement with a light scatter angle set to backscatter 175 °C. Prior to measurement, each sample was diluted with PBS to obtain a suitable particle concentration for analysis and gently mixed to ensure homogeneity. Samples that contained visible aggregates were filtered through a 0.22 μm syringe filter to prevent interference with the measurement.

3.7 Purification of Nanoparticles

The resulting nanoparticles were purified using centrifugal filter units (Sartorius Vivaspin 2) with a molecular weight cut-off (MWCO) of 100 kDa to remove unencapsulated molecules and DMSO. In this setup, intact RBCVs (+/- cRGD) were retained above the membrane, whereas free AA (~176 Da) are far below the MWCO and therefore pass into the filtrate. Prior to use, all filters were pre-rinsed with PBS by centrifugation at 4000 rpm for 5 minutes at 4 °C to eliminate any manufacturing residues that could interfere with the samples. The RBCV solution was then added into the filter units and centrifuged under the same conditions for 40 minutes. The filtrate, which contained excess PBS, DMSO, and non-assembled components, was eliminated after each spin. To avoid membrane surface appearing discoloured or partially clogged, the retained suspension was gently pipetted up and down to resuspend them. A second wash with PBS was performed to enhance purity. Finally, the concentrated nanoparticle was collected by reverse spin at same conditions. The concentrated nanoparticles were collected carefully at the tap and stored at 4 °C for subsequent analysis.

3.8 Membrane distraction

To quantify the amount of AA encapsulated within RBCVs, the purified vesicles were lysed using Triton X-100 (TX-100). TX-100 was added directly into the filter unit to a final concentration below 1% (v/v), followed by gentle mixing and incubation at 37 °C for 10 minutes to fully disrupt the vesicle membranes. This released all encapsulated contents into solution. After lysis, the samples were centrifuged at 4000 rpm for 40 minutes at 4 °C, and the filtrate containing the released AA was collected for quantification. Encapsulation efficiency (EE%) could be calculated based on the AA calibration curve.

3.9 Halo Clot Assay

The halo clot assay was used to assess the thrombolytic activity of the nanoparticle formulations. A clotting buffer containing 0.5M Tris-HCl, 5mM NaCl and 0.37M CaCl_2 were prepared and adjusted to pH 7.42. 5 μL of 1 μM Thrombin was added immediately before the clotting process.

In each well of a 96-well plate, 15 μL of whole human blood and 5 μL of clotting buffer were added on opposite sides and gently mixed to initiate clot formation. The plate was incubated at 37 °C for 30 minutes to allow stable clots to develop.

Following clot formation, 80 μL of each test sample was added to the centre of the clot. Positive control included non-clotted whole blood, while negative control was clotted blood with PBS only. The plate was then placed in a plate reader for 90 minutes, which recorded clot degradation over time

by measuring halo formation. The overall workflow is shown in following Figure 3.9.1.

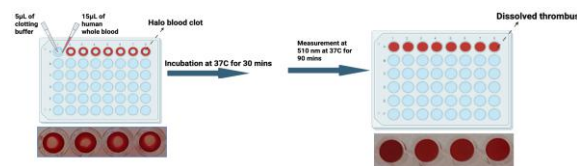


Figure 3.9.1 Halo Clot Assay workflow

3.10 Calculations

The encapsulation efficiency (EE%) of the nanoparticles was determined by quantifying the amount of unencapsulated drug in the filtrate after purification. The concentration of free drug was measured using UV–Vis spectrophotometry, and the total drug amount was defined as the quantity initially added during nanoparticle fabrication. Encapsulation efficiency was calculated using the following equation:

$$EE\% = \frac{W_{total} - W_{free}}{W_{total}} \times 100\% \quad (1)$$

Where W_{total} represents the total amount of drug used in the formulation and W_{free} denotes the amount of unencapsulated drug detected in the filtrate. All measurements were performed in triplicate, and the average EE% was reported with standard deviation.

For the halo clot assay, thrombolytic activity was assessed by quantifying the area or diameter of the halo zone formed after sample addition. Images obtained from the plate reader were analysed using ImageJ software. The clot degradation efficiency (CD%) was expressed as the ratio between the halo area of the treated sample and that of the positive control, calculated according to the following formula:

$$CD\% = \frac{A_{sample} - A_{negative}}{A_{positive} - A_{negative}} \times 100\% \quad (2)$$

Where A_{sample} is the measured absorbance of the test sample, $A_{positive}$ corresponds to complete clot degradation (positive control), and $A_{negative}$ represents the untreated clot. Data were analysed using one-way ANOVA, and results were expressed as mean $\pm$ standard deviation (SD). A p-value < 0.05 was considered statistically significant.

4. Results and discussion

4.1 Validation of the AA–DCPIP Quantification Framework

As mentioned above in methodology, a calibration curve was first established to accurately quantify AA. As shown in Figure 4.1.2, initial optimisation indicated that 190.3 $\mu\text{g/mL}$ of DCPIP yielded a strong negative correlation with an increasing ascorbic acid concentration. In contrast, when DCPIP was diluted further, the reaction became too fast, and the test samples fell below the detection limit before the assay was read (figure 4.1.1). As a result, DCPIP (190.3 $\mu\text{g/mL}$) was chosen as the

analytical indicator concentration for all subsequent AA calibration curves. Figure 4.1.2. shows the final calibration curve for AA against DCPIP.

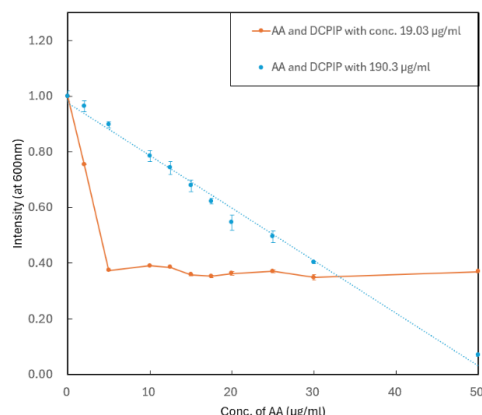


Figure 4.1.1 Calibration curve of AA reacted with 190.3 $\mu\text{g/mL}$ dilution DCPIP

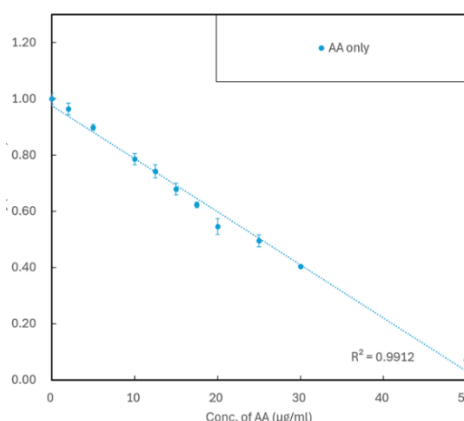


Figure 4.1.2 Calibration curve of AA reacted with 1/10 dilution DCPIP

Since streptokinase (SK) was planned to be co-encapsulated with AA during final vesicle preparation, its potential impact on AA detection was evaluated. Calibration curves developed in the presence of SK (concentration-dependent) shown in Figure 4.1.3 demonstrated no measurable deviation in slope or intercept compared to those containing AA alone. These results indicate that SK neither reacts with AA or DCPIP nor alters the optical properties of the assay, confirming the suitability of this detection method for samples containing SK.

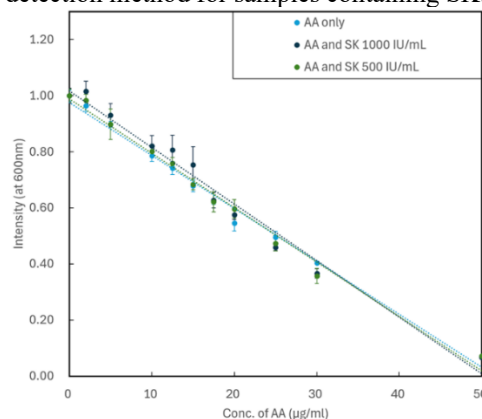


Figure 4.1.3 Impact of SK on AA quantification

With the AA–DCPIP assay validated in terms of both indicator concentration and SK interference,

initial experiments examined whether unloaded RBCVs affected AA quantification. As shown in Figure 4.1.4, higher concentrations of RBCVs introduced variability into the AA–DCPIP titration curve, while increasing dilution (e.g., to 1/10) gradually restored linearity and produced a response closer to the control.

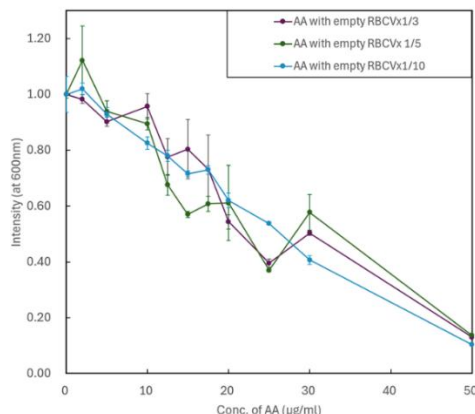


Figure 4.1.4 Impact of RBCV on AA quantification

However, such high levels of dilution could not be applied in the final loading assay since the expected AA concentration itself is low. Diluting the RBCV samples to this extent would also dilute the encapsulated AA, potentially lowering its concentration beyond the detectable limit of the assay. This meant that RBCV dilution was not a practical solution for minimising matrix effects. Since residual haemoglobin is commonly present in final RBCV solutions, a secondary calibration curve of AA in the presence of haemoglobin was derived from the same concentration of ghost cells that are used in the preparation of RBCVs.

As shown in Figure 4.1.5, haemoglobin caused a slight upward shift in absorbance, but the linearity and slope of the calibration curve were preserved ($R^2 > 0.99$), giving equation (3) specified. Importantly, the haemoglobin interference remained consistent across experiments because each measurement used ghost cells at the same concentration and volume, ensuring a reproducible baseline shift that could be reliably accounted for.

$$A = -0.0145C_{AA} + 1.0117 \quad (3)$$

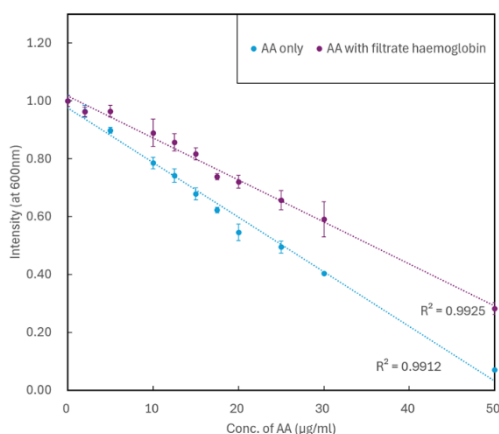


Figure 4.1.5 Impact of haemoglobin on AA quantification

Overall, these results indicate that AA–DCPIP quantification remains reliable in the presence of physiologically relevant haemoglobin levels, supporting the use of this assay for AA loading measurements throughout the study.

4.2 Preparation of Ghost cells and Characterization of Red blood cell vesicles

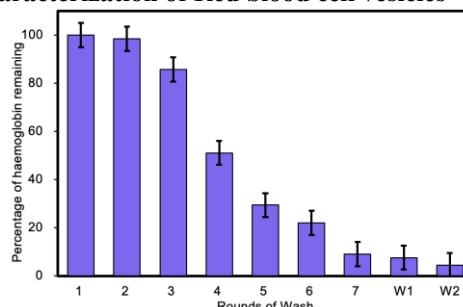


Figure 4.2.1 Haemoglobin remain after each wash

The removal of haemoglobin during ghost cell preparation was evaluated by testing the supernatant across cycles. The results are shown in Figure 4.2.1. After the first three rounds of washing, a high proportion of haemoglobin remained within the samples, with values above 80%, indicating that the initial washes primarily removed extracellular components but did not yet achieve substantial lysis or haemoglobin clearance. A significant decrease was shown in the Figure 4.2.1 after the fourth wash onward. Haemoglobin levels dropped significantly to approximately 50%, which suggesting that repeated hypotonic lysis wash and centrifugation effectively remove haemoglobin in the process.

Further reductions were observed in subsequent washes, with the fifth and sixth cycles reducing the remaining haemoglobin to approximately 30% and 20% respectively. By the seventh wash, the haemoglobin content had decreased to below 10%. Further washing with PBS eliminated all the other released haemoglobin.

This gradual reduction demonstrates that the repeated hypotonic wash cycles effectively remove haemoglobin from red blood cells, preparing them for downstream processing.

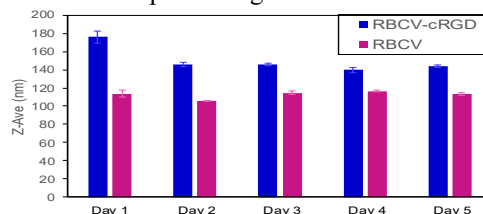


Figure 4.2.2 Average diameters of Nanoparticles

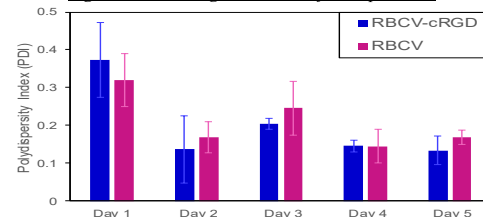


Figure 4.2.3 Average PDI of Nanoparticles

The DLS measurements of nanoparticles over five days are summarised in Figures 4.2.2 and 4.2.3. Throughout the measurement period, both formulations remained within the nanoscale range. RBCV-cRGD nanoparticles consistently showed a larger hydrodynamic diameter than the non-targeted RBCVs on all days, typically around 140–170 nm compared with approximately 100–120 nm for RBCVs. This size increase is expected as cRGD is present on a PEG polymer which extends from the vesicle membrane, resulting in an increased hydrodynamic radius. There was no clear tendency for particle size to decrease over time, indicating that the nanoparticles remained stable and did not undergo structural collapse or shrinkage.

The polydispersity index (PDI) data indicate a similar pattern of stability, with a noticeable exception on Day 1 that both formulations displayed relatively high PDI values, particularly RBCV-cRGD, implying a broader size distribution. This was later attributed to the presence of dust or other small particulates in the cuvette rather than intrinsic instability of the nanoparticles. From Day 2 onwards, the samples were passed through a filter before DLS analysis. Following this additional cleaning step, the PDI values for both formulations decreased and were still below 0.3, reflecting a reasonably uniform dispersion over Days 2–5. Overall, these results suggest that the RBCVs and RBCV-cRGD's formulations are structurally stable over several days.

4.3 Loading Efficiency of Ascorbic Acid (AA)

AA loaded into RBCVs was assessed following purification of all samples through the same ultrafiltration protocol, ensuring that only vesicle-associated AA remained. The absorbance values obtained were used to calculate the loaded AA through the derived calibration curve and the resulting encapsulation efficiency (EE) listed in Table 4.3.1 below.

Table 4.3.1 Encapsulation results

Conc. Of AA loading into the RBCV	Notes	Absorbance AVE	Conc. Of AA encapsulated	EE (%)
50.00	Without cRGD dissolved in DMSO	0.89	15.43	30.85
100.00	Without cRGD dissolved in DMSO	0.77	22.98	22.98
500.00	Without cRGD dissolved in DMSO	0.63	33.18	6.64
500.00	Without cRGD dissolved in DMSO & incubation	1.13	3.45	0.69
500.00	With cRGD dissolved in DMSO	1.24	12.91	2.58

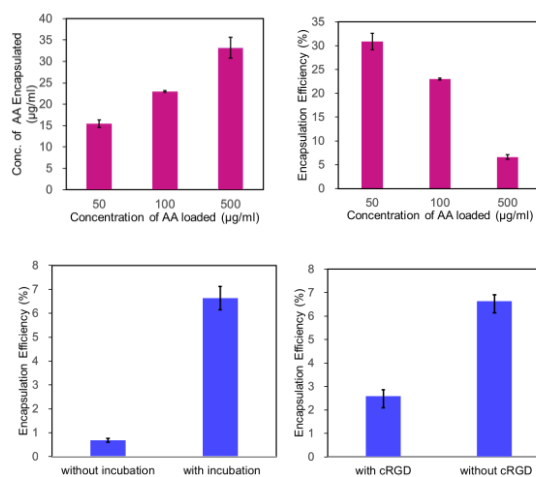


Figure 4.3.1 Encapsulated AA & EE% depends on different loading conc. And conditions

As shown in Figure 4.3.1, increasing the AA concentration from 50 to 100 µg/mL resulted in a proportional increase in the amount of AA retained within RBCVs, consistent with passive diffusion-driven loading. However, when the loading concentration was increased to 500 µg/mL, the encapsulation efficiency did not appear to increase further. While the maximum concentration of encapsulated AA was around ~30 µg/mL, the calculated encapsulation efficiency was only about 7% against the initial loading concentration. This behaviour suggests that the vesicles reach a saturation threshold beyond which additional AA cannot be accommodated and is subsequently removed during washing. High solute concentrations may also impose osmotic stress, increase leakage or destabilise the vesicle membrane, which further reduces retention.



Figure 4.3.2 Halo clot assay with free AA added

In this assay, the wells on the right correspond to AA concentrations above 100 µg/mL, where AA was added in its free (non-encapsulated) form. These wells show a clear brownish colour, indicating oxidation of haemoglobin and damage to the membrane. In contrast, the wells on the left were prepared with lower AA concentrations, which did not cause visible oxidation. Indicating that excess free AA can induce oxidative damage in the system, confirming that AA concentrations above 100 µg/mL should be avoided during formulation.

The influence of surface functionalisation was also examined. cRGD-modified RBCVs did not show any improvement in AA loading and in fact exhibited lower encapsulation than unmodified vesicles. This observation aligns with expectations, as cRGD acts primarily as an external targeting ligand and is not intended to enhance membrane permeability, in addition, liposomes functionalized with cRGD had an encapsulation efficiency (EE) for co-loaded small-molecule drugs greater than 85%,

essentially unchanged compared to non-modified liposomes. Peptide decoration may also introduce mild steric or packing effects on the membrane surface, limiting the passive entry of small molecules such as AA. (Wu & Zeng, 1 June 2024). In our system, low concentrations of DMSO had negligible impact on the encapsulation efficiency of ascorbic acid, whereas a significant increase in EE was observed for streptokinase in the presence of DMSO.

4.4 Functional Assessment: Thrombolytic Performance (Halo clot)

To evaluate the individual and combined effects of AA, streptokinase (SK), and RBCV-based delivery on thrombolytic performance, a halo clot assay was performed using a comprehensive set of control and test formulations.

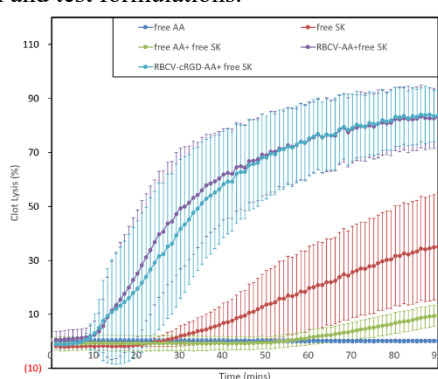


Figure 4.4.1 Percentage degradation of clots over 90 minutes trial 2

Figure 4.4.1 compares several samples, including free AA (12.5 $\mu\text{g/mL}$), free SK (100 IU/mL), free AA (12.5 $\mu\text{g/mL}$) + free SK (100 IU/mL) mixtures, RBCV AA (33.18 $\mu\text{g/mL}$) + free SK (100 IU/mL) and RBCV-cRGD AA (12.91 $\mu\text{g/mL}$) + free SK (100 IU/mL).

AA alone showed almost no thrombolytic activity, remaining close to baseline throughout the assay. In contrast, AA-loaded RBCV formulations showed a much steeper increase in clot degradation. Both RBCV-loaded AA and RBCV + cRGD-loaded AA groups reached around 80–100% degradation, with the targeted version showing a slightly faster early rise but similar final performance compared to the non-targeted RBCV.

The AA+SK mixture performed poorly compared to SK only, but encapsulated AA of the same concentration with free SK significantly enhanced clot lysis indicating the role of the nanocarrier.

From results above, it was clear that free AA and the free AA+SK mixture showed minimal thrombolytic activity, so these two groups were excluded from the subsequent experiments.

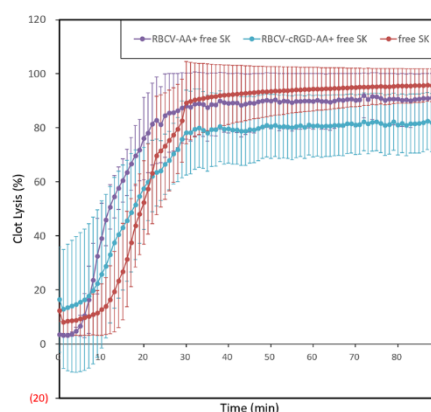


Figure 4.4.2 Percentage degradation of clots over 90 minutes

The graph above (Figure 4.4.2) compares thrombolysis over time for three tested groups: free SK (100 IU/mL), RBCV AA (33.18 $\mu\text{g/mL}$) + free SK (100 IU/mL) and RBCV-cRGD AA (12.91 $\mu\text{g/mL}$) + free SK (100 IU/mL). All groups showed minimal activity during the first few minutes, followed by a gradual increase in clot degradation.

Around 30 minutes, all groups had already reached most of their final degradation level, with only gradual increases afterward, indicating that most of the clot dissolution occurred early in the assay. Overall, the results suggest that AA-loaded RBCV formulations enhance the thrombolytic effect compared to free SK, although cRGD modification in this formulation did not enhance performance.

In comparing these two graphs for thrombolysis, the overall magnitude and rate of clot degradation vary significantly between the two experiments. The same formulations were compared, but it is evident that both curves have obvious differences. Different blood donors provide the most likely reason for this difference. Donor-to-donor variability, differences in haematocrit, fibrin structure, clot density, platelet activation levels, and intrinsic coagulation properties can powerfully contribute to a clot's formation and its plasticity in the face of thrombolytic agents. Consequently, even identical formulations can produce different kinetic profiles in experiments. This demonstrates the inherent biological variability of whole-blood assays, as well as the necessity of carrying out experiments with multiple donors, or carefully normalising data for comparison between separate datasets.

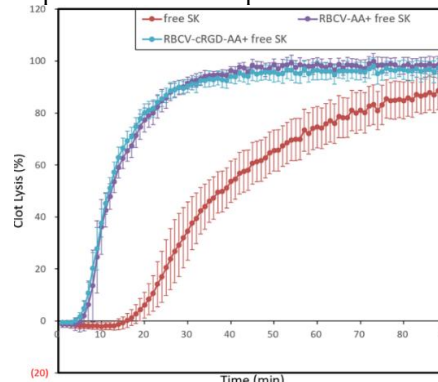


Figure 4.4.3 Percentage degradation of clots over 90 minutes trial 3

The halo clot assay was repeated to evaluate the performance of AA-RBCVs. The results are shown in Figure 4.4.3. In this repeat experiment, all formulations again showed a rapid rise in clot lysis within the first 20–30 minutes. The RBCV-based formulations displayed a noticeably faster increase during the initial phase compared to free SK, reaching high degradation levels much earlier. Both the RBCV-loaded AA and the RBCV-cRGD formulations achieved final degradation values of over 90%, while free SK remained substantially lower, around 70–80%.

The overall trend and final values closely resembled those seen in the first dataset, supporting the idea that the earlier discrepancy was mainly caused by donor-to-donor biological variability rather than the formulations themselves. Since whole blood assays are sensitive to individual differences in clot density, platelet activation and fibrin structure, it is expected that performance may vary when different donors are used. The repeat assay confirms that using the same donor restores the expected pattern, and reinforces that the RBCV-based systems produce a faster onset of thrombolysis and a higher overall effect compared with free SK.

5 Conclusion

This work explored red blood cell-derived vesicles (RBCVs) as nanocarriers for streptokinase (SK) with co-loaded ascorbic acid AA as an adjuvant to enhance thrombolytic performance.

First, an AA-DCPIP colorimetric assay was successfully optimised for AA quantification in the presence of both SK and RBCVs. At 190.3 µg/mL DCPIP concentration, the assay remained linear and reproducible despite minor baseline shifts caused by residual haemoglobin, enabling reliable evaluation of AA loading across formulations.

Second, AA encapsulation within RBCVs was found to be limited by membrane permeability and saturation effects. While AA loaded increased at lower loading concentrations, a clear plateau was observed at higher concentrations, leading to reduced encapsulation efficiency and indicating a finite loading capacity of the vesicle membrane. Additionally, surface functionalisation with cyclic RGD (cRGD) did not enhance the encapsulation of the small molecule AA, which is consistent with cRGD functioning primarily as a targeting ligand rather than a permeability modifier.

Finally, functional halo clot assays demonstrated that AA-loaded RBCVs, with or without cRGD, in the presence of SK achieved enhanced clot lysis compared with free SK or free AA. These results suggest a synergistic effect between AA and SK when delivered via RBCVs, supporting RBCVs as a promising platform for improving thrombolytic efficacy.

6 Outlook

Looking forward, there are several future directions can be explored to further strengthen this work. At first, the thrombolytic performance of the formulations should be evaluated in clots with different platelet content, as ascorbic acid is proposed to primarily interfere with platelet–platelet interactions. This would enable a more systematic assessment of how clot composition influences treatment efficacy.

Second, more detailed mechanistic studies focusing on platelet interactions are required. Comparing the binding behaviour of RBCVs and cRGD-functionalised RBCVs to activated platelets would help clarify whether the targeting ligand significantly alters clot–nanoparticle interactions.

Finally, the antioxidant function of AA-loaded RBCVs could be investigated using cell-based or tissue-level oxidative stress models. Such studies would allow evaluation of whether the antioxidant cargo provides additional protective effects during or after thrombolysis.

7 Acknowledgement

We wish to express our profound gratitude to Prof. Chen for his consistent guidance and support over the duration of this project. We are also deeply grateful to PhD student Diya Agrawal for her committed supervision and indispensable help in experimental procedures, without which this work would not have been possible. Furthermore, we gratefully acknowledge the kindness and assistance of all members of Prof. Chen's research group.

References

- Agnieszka, G., & Skrzydlewska, E. (2022 Oct 7). *Antioxidative and Anti-Inflammatory Activity of Ascorbic Acid*.
- Aguilar, K., Garvín, A., Ibarz, A., & Augusto, P. (2017). Ascorbic acid stability in fruit juices during thermosonication. *Ultrasonics Sonochemistry*, 37, 375–381.
- Alkarithi, G., Duval, C., Yu, S., Macrae, F., & Ariëns, R. (2021). Thrombus Structural Composition in Cardiovascular Disease. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 41(9), 2370–2383. doi:10.1161/ATVBAHA.120.315754
- Arauna, D., Navarrete, S., Cecilia, A., Wehinger, S., Pizarro-Mena, R., Palomo, I., & Fuentes, E. (2024). Understanding the Role of Oxidative Stress in Platelet Alterations and Thrombosis Risk among Frail Older Adults. *Biomedicine*, 12(9), 2004. doi:10.3390/biomedicine12092004
- C N, C. (1998). *The vascular endothelium: a regulator of hemostasis and thrombosis*. *Blood*, 91(10), 3527–3561. (Vol. 2).

- Cines, D. B. (1998). *The vascular endothelium: a regulator of hemostasis and thrombosis. Blood*, 91(10), 3527–3561.
- Collen D, L. H. (1995). Molecular basis of fibrinolysis, as relevant for thrombolytic therapy. *Thrombosis and Haemostasis*, 167-171.
- Collen, D. (2001). *The plasminogen (fibrinolytic) system. Thrombosis and Haemostasis*, 86(1), 266–274.
- D, C., & H, R. L. (1995). Molecular basis of fibrinolysis, as relevant for thrombolytic therapy. *Thrombosis and Haemostasis*, 74(1), 167-171.
- Davi, G., & Patrono, C. (2007). *Platelet activation and atherothrombosis* NEJM, 357(24), 2482–2494. (Vol. 357).
- Désiré, C. (2001). *The plasminogen (fibrinolytic) system. Thrombosis and Haemostasis*, 86(1), 266–274 (Vol. 82).
- Ellis, G. R., Anderson, R. A., Chirkov, Y. Y., Morris-Thurgood, J., Jackson, S. K., Lewis, M. J., . . . Frenneaux, M. P. (2001). Acute Effects of Vitamin C on Platelet Responsiveness to Nitric Oxide Donors and Endothelial Function in Patients with Chronic Heart Failure. *Journal of Cardiovascular Pharmacology*, 37(5), 564-570. doi:10.1097/00005344-200105000-00008
- Freedman, J. E. (2008). Oxidative Stress and Platelets. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 28(3), 405-412. doi:10.1161/ATVBAHA.107.159178
- Henry, N. (2022). Cell Membrane-Derived Nanoscale Systems for Drug Delivery Applications.
- Jilani, T. N., & Siddiqui, A. H. (2023, 2 20). *Tissue plasminogen activator*. Retrieved from StatPearls Publishing: <https://www.ncbi.nlm.nih.gov/books/NBK556123/>
- Li, Z., Bi, R., Sun, S., Chen, S., Chen, J., Hu, B., & Jin, H. (2022). The Role of Oxidative Stress in Acute Ischemic Stroke-Related Thrombosis. *Oxidative Medicine and Cellular Longevity*, 2022:1–19. doi:10.1155/2022/8418820
- Liu, C.-H., Rethi, L., Weng, P.-W., Nguyen, H., & Chuang, A.-Y. (2024). *Cutting-edge advances in nano/biomedicine: A review on transforming thrombolytic therapy*.
- Liu, D., Pei, D., Hu, H., Gu, G., & Cui, W. (2020). Effects and Mechanisms of Vitamin C Post-Conditioning on Platelet Activation after Hypoxia/Reoxygenation. *Transfusion Medicine and Hemotherapy*, 47(2), 110-118. doi:10.1159/000500492
- Nguyen, P., Jayasinghe, M., Le, A., Peng, B., & Le, M. (2023). Advances in Drug Delivery Systems Based on Red Blood Cells and Their Membrane-Derived Nanoparticles. *ACS Nano*, 17(6), 5187-5280. doi:10.1021/acsnano.2c11965
- Raghavan, S., Sharma, P., & Dikshit, M. (2003). Role of ascorbic acid in the modulation of inhibition of platelet aggregation by polymorphonuclear leukocytes. *Thrombosis Research*, 110(2-3), 117-126. doi:10.1016/S0049-3848(03)00312-8
- Raskob, A. M. (2016). Global Burden of Thrombosis: Epidemiologic Aspects. *Circulation Research*, 1340-1347.
- Subhradip Mukhopadhyay, S. M. (2019). *Fibrinolysis and Inflammation in Venous Thrombus Resolution*.
- T V Hrynenko I, I. M.-W. (2002). *Degradation of streptokinase and the catalytic properties of the plasmin-streptokinase complex*.
- Tomkins, A. J., Schleicher, N., Murtha, L., Kaps, M., Levi, C. R., Nedelmann, M., & Spratt, N. J. (2015). Platelet rich clots are resistant to lysis by thrombolytic therapy in a rat model of embolic stroke. *Experimental & Translational Stroke Medicine*, 7(2), 1-9. doi:10.1186/s13231-014-0014-y
- Tyml, K. (2017). Vitamin C and Microvascular Dysfunction in Systemic Inflammation. *Antioxidants*, 6(3), Article 49. doi:10.3390/antiox6030049
- Wendelboe, A. M., & Raskob, E. G. (2016). Global Burden of Thrombosis: Epidemiologic Aspects. *Circulation Research*, 118(9), 1340-1347.
- White, H. (1989). Comparison of Tissue Plasminogen Activator and Streptokinase in the Management of Acute Myocardial Infarction. *Chest*, 95(5), 265S-269S. doi:10.1378/chest.95.5.265S
- Wu, Y., & Zeng, C. (1 June 2024). *Tumor-Targeted cRGD-Coated Liposomes Encapsulating Optimized Synergistic Cepharanthine and IR783 for Chemotherapy and Photothermal Therapy*.
- Xia, Q., Zhang, Y., Li, Z., Hou, X., & Feng, N. (2019). Red blood cell membrane-camouflaged nanoparticles: A novel drug delivery system for antitumor application. *Acta Pharmaceutica Sinica B*, 9(4), 675-689.

Modelling of 3D Solute Transport in Disordered Porous Media

Erin Richards and Eszter Gyuris

Department of Chemical Engineering, Imperial College London, U.K.

Supervisor: Dr. Ronny Pini

December 11, 2025

Abstract

Modelling of fluid in porous rock is made difficult by the complex pores structure and pore connectedness. This study presents a 3 dimensional model of transport and flow within porous media by accounting for both macro- and micro porosity, as well as the spatial heterogeneity of the pore structure. The model is implemented using MODFLOW-2005 and validated against high-resolution PET imaging data. A dual-porosity approach is implemented, distinguishing mobile and immobile pore domains with intragranular mass transfer, k , to capture the effects between the two. Permeability heterogeneity is represented using both fractal geometry and scaling factor correlations. Model predictions are validated against high-resolution PET-imaged tracer experiments, comparing breakthrough curves, spatial moments, and three-dimensional concentration distributions. Results show that fractal geometry models capture general transport trends but fail to reproduce temporal spread and detailed plume structure, while scaling factor heterogeneity with an exponent, $B = 3$, and $k = 0.0001\text{--}0.001\text{ s}^{-1}$ closely matches experimental observations. The study accurately predicts 3 dimensional transport in heterogeneous porous rocks and is a key stepping stone for modelling porous systems in the future.

1 Background

Modelling transport in porous media is key in the engineering of clean water supplies and CO₂ storage. In both cases, solving these challenges depends on how dissolved material moves through porous media, and how accurately and reliably we can predict this transport.

Groundwater is an essential resource, providing nearly half of the world's drinking water [1], with many dry regions depending on this source. With increasing pressures from population growth and climate change, the stability of this resource is becoming increasingly uncertain due to over exploitation and pollution. Contaminants introduced by agricultural and industrial activity, such as pesticides, hydrocarbons and chlorinated solvents [2] are difficult to remove from aquifers via remediation. This is due to the complexity in flow patterns including advection, dispersion, mass transfer between different pores and strong heterogeneity in the natural structure of porosity and permeability. Any groundwater remediation strategy will therefore rely heavily on an accurate and validated transport model, able to replicate this solute behaviour three dimensionally and to represent the effects of geological heterogeneity in a robust way.

A similar issue exists concerning geological CO₂ storage, an important solution in the long term management of harmful carbon emissions. The effectiveness of optimal injection sites for storage is dependent on how the CO₂ solution moves through the pores. At the reservoir scale, spatial variations in porosity and permeability, and the presence of high and low permeability channels strongly influence flow

pathways and trapping mechanisms[3]. This heterogeneity involves regions of lower conductivity increasing the residence time for dissolved carbon to become a solid phase[4]. Additionally, heterogeneity affects pressure fluctuations and risk of leakage. Due to this importance of the underlying geological structure, the accurate locating and monitoring of optimal injection sites relies heavily on precise and reliable three dimensional transport models. When modelling flow and transport through porous rocks at a laboratory scale, microporosity is often neglected to focus on the bulk of the concentration. Many studies have shown that the neglect of microporosity in scale up has a significant effect on the error [5], highlighting the need for more detailed laboratory scale modelling to minimise leakage and any prevent disasters.

2 Introduction

This project investigates the improvement of predictive three-dimensional modelling of transport through porous media, to capture the effects of micropores. Our benchmark material, Ketton limestone, has a highly varied pore structure. The structure contains macro pores ($> 50\text{ [nm]}$), which are the spaces between the grains often interconnected, allowing for flow, in this study we will refer to this as the mobile domain. Micro pores ($< 1\text{ [nm]}$) which are the small cavities within the grain, are also present. In micropores, fluid is often isolated and stagnant, we will refer to this the immobile domain. These features make predicting solute transport within porous rock difficult, as the velocity field becomes highly varied. The more disordered and heterogeneous the porosity field, the

harder it is to predict the shape and dispersion of a concentration plume within the core.

A major knowledge gap is the lack of evidence confirming the most accurate method to describe the heterogeneity that is responsible for spatial variance and mixing in porous rock systems. Traditionally, porosity correlations such as Kozeny–Carman relationship, fractal geometry [6], or multi-phase fluid mapping methods have been used to estimate permeability [7]. However, there is little confirmation on any of these approaches.

In past research, analysis largely focused on the outlet values of tracer pulse experiments, mapping the outlet concentration vs time, or breakthrough curves (BTC). From these studies, it has been determined that the Advection Dispersion Equation(ADE) is the most suitable model [8]. This model is commonly applied only in one dimension, and can only give information about the average dispersion in the direction of the flow, but not the behaviour of the internal plume. As many factors, such as permeability and porosity, can affect the breakthrough curve, the internal concentration profile is crucial to validate the ADE model. Furthermore, when 3 dimensional modelling has been conducted there has been little experimental data to validate the model [9]. This project aims to validate our predictions using a high resolution PET imaging dataset taken from a previous study [10] [11].

The aim of this study is to accurately model the flow and transport of a tracer pulse experiment in 3 dimensions using MODFLOW-2005. We do this by introducing micro porosity and permeability correlations, where parameters such as the permeability model, longitudinal dispersivity and mass transfer coefficient are used to fit the model to the experimental data. Breakthrough curves and spatial moments are used to quantitatively compare experimental values, while 3 dimensional concentration maps qualitatively validate our model.

This study presents a 3D, experimentally validated model that captures detailed pore-scale heterogeneity and transport, bridging the gap between simple 1D models and realistic predictions for large-scale environmental applications.

3 Imaging Dataset

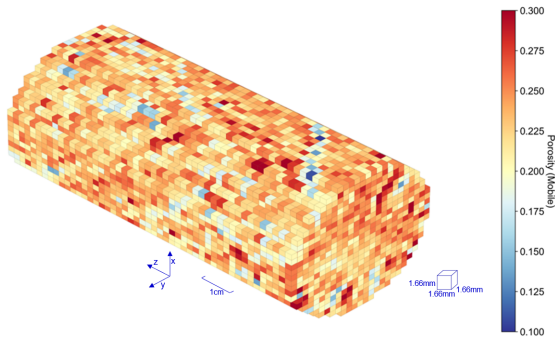


Figure 1: Total porosity field of Ketton limestone sample cylinder generated in Python from XCT imaging [11]. (length: 10 [cm], diameter: 5 [cm], voxel size $2 \times 2 \times 2$ [mm³])

This study uses an experimental dataset from the Stanford digital respiratory [11]. The respiratory includes two datasets. First is from X-ray Computed Tomography (XCT) data from the core in conditions of saturated air and water to generate three-dimensional porosity, and scaling factor maps (Figure 1). With voxel sizes $2 \times 2 \times 2$ [mm³] the data is stored in matrix form.

The second dataset extracted was generated from a pulse tracer experiment at 4 ml/min using radio tracer (18F-FDG) in a core flooding apparatus[8]. The tracer is a [18F] Fluorodeoxyglucose radiotracer with a halflife of $t_{1/2} = 109.7$ [minutes]. This is dissolved in a dilute aqueous KCl carrier fluid at 0-2.3 [wt%] KCl and a tracer concentration of 0.278 [mCi·ml⁻¹].

During the experiment the rock sample is divided into a grid with voxel dimensions of $0.776 \times 0.776 \times 0.796$ [mm³], and Positron Emission Tomography (PET) is used to obtain concentration values for each voxel. The data is stored in matrix form, and the voxel values show decay-corrected radioactivity intensity. Both datasets are subsequently translated and visualised in a MATLAB environment.

3.1 Materials

The extracted experimental data is sampled from a Ketton limestone sample[11] of length 10[cm] and diameter 5[cm], with smooth, spherical grains of diameter $d_p = 0.06$ [cm]. The sample is mono-mineralic with 99[%] CaCO₃, with a measured average porosity, $\phi_t = 0.234[-]$, measured mean immobile porosity, $\phi_{im} = 0.085[-]$ and mean permeability, $K_{mean} = 1925$ [mD] [8].

4 Modelling

MODFLOW solves partial differential equations in 3 dimensions to model flow in saturated porous media (3). The model assumes a single-phase fluid with constant density, viscosity, and temperature. The flow is considered steady state and laminar, and Darcy’s Law applies (2). The model also assumes the principal directions of hydraulic conductivity are Cartesian and do not vary within the system [12].

$$Q = -K_{mean} \cdot A \frac{\Delta h}{\Delta L} \quad (1)$$

$$Q = -\mathbf{K} \cdot A \nabla h \quad (2)$$

where Q is the volumetric flow rate [$cm^3 \cdot s^{-1}$], $\mathbf{K}$ is the hydraulic conductivity vector [$cm \cdot s^{-1}$], A is the cross sectional area [cm^2], h is the hydraulic head [cm] and ΔL is the flow path length [cm].

$$S_s \frac{dh}{dt} + Q = \nabla \cdot (\mathbf{K} \nabla h) \quad (3)$$

$$0 = \frac{\partial}{\partial x} (K_x \frac{\partial h}{\partial x}) + \frac{\partial}{\partial y} (K_y \frac{\partial h}{\partial y}) + \frac{\partial}{\partial z} (K_z \frac{\partial h}{\partial z})$$

where S_s is the storage coefficient [cm^{-1}]. Under steady state flow with negligible pressure variations $S_s \frac{dh}{dt} \approx 0$, and $Q = 0$

everywhere as there is no generation or consumption.

MODFLOW then uses the finite differences scheme to solve the groundwater flow equation by dividing up the domain into a grid of 3 dimensional voxels. The dimensions of the grid are specified and each voxel is allocated a row, column and layer. Equation (3) is applied to each voxel to form a systems of equations stored in matrix form. This equation matrix is then solved iteratively to acquire the velocity field.

4.1 Modified Advection Dispersion Equation (ADE)

MODFLOW is paired with MT3DMS which uses the velocity field generated to solve the 3 dimensional ADE. It is assumed that there is no reaction or tracer decay and that there are no sources or sinks within the grid, only at the boundary conditions.

$$\phi_{mob} \frac{\partial C}{\partial t} + \phi_{im} \frac{\partial C_{im}}{\partial t} = \nabla \cdot (\phi_{mob} D_l \nabla C) - \nabla(qC) + k(C - C_{im}) \quad (4)$$

where C and C_{im} are the concentrations in the mobile and immobile pore space respectively [-], k is the intragranular mass transfer coefficient [s^{-1}]. q is the Darcy's flux or specific discharge ($\mathbf{q} = \frac{\mathbf{Q}}{A} = -\mathbf{K} \nabla h$), and D_l is the longitudinal dispersion coefficient:

$$D_l = D_{eff} + \alpha_l v \quad (5)$$

where D_{eff} is the effective diffusion coefficient, which is taken as the bulk diffusion coefficient $\mathcal{D} = 1e^{-5} [cm^2 \cdot s^{-1}]$, α_l is the longitudinal dispersivity and v is the pore interstitial fluid velocity [$cm \cdot s^{-1}$],

$$v = \frac{u}{\phi_{mob}} \quad (6)$$

where u is the superficial velocity [$cm \cdot s^{-1}$]. It is assumed that there is negligible transverse dispersivity, as it is a factor of ten smaller than longitudinal dispersivity (Ketton limestone: $\alpha_l = 0.16[cm]$ [13] vs. $\alpha_t = 0.013[cm]$ [14]).

4.2 Solution Procedure

The initial condition sets the concentration throughout the domain to zero:

$$C(z, 0) = 0, \quad 0 \leq z \leq L, \quad (7)$$

while the inlet uses the Danckwerts boundary condition:

$$-D \frac{\partial C}{\partial z} \Big|_{z=0} + u C(0, t) = u C_{in}, \quad (8)$$

and the outlet uses a zero-gradient condition:

$$\frac{\partial C}{\partial z} \Big|_{z=L} = 0. \quad (9)$$

This setup ensures advection–dispersion behavior without artificial back-diffusion, with all concentrations starting from zero. The code completes a maximum of 300 iterations and has a convergence criteria, or tolerance of $1e-4$.

4.3 Porosity Fields

Porosity fields can be characterised in two ways: the space in the mobile (ϕ_{mob}) and immobile (ϕ_{im}) domains. The model is tested on two porosity fields: synthetic fields created to assess its physical validity, and the experimentally measured porosity field of Ketton limestone obtained from the data repository [11].

The synthetic mobile porosity is attained by generating a 3 dimensional randomly distributed porosity field in the same dimensions as the experimental sample. The synthetic model is designed around the mean porosity of Ketton limestone calculated from experimental data, ($\phi_t = 0.23$) [11], with a standard deviation of 0.02 and immobile porosity, $\phi_{im} = 0.085$ [8].

The experimental porosity fields are generated from the XCT scans. This data is the total space available or total porosity. It is assumed that the immobile field is constant value of $\phi_{im} = 0.085$. To relate this to the mobile porosity fields, the following correlation is made:

$$\begin{aligned} \phi_{tot} &= \phi_{mob} + (1 - \phi_{mob})\phi_{im} \\ \phi_{mob} &= \frac{\phi_{tot} - \phi_{im}}{1 - \phi_{im}} \end{aligned} \quad (10)$$

4.4 Hydraulic Conductivity Fields

Hydraulic conductivity, K [$cm \cdot s^{-1}$], or permeability, k [mD], provides the primary means of introducing heterogeneity into the model.

$$\mathbf{K} = \mathbf{k} \cdot \frac{\rho \cdot g}{\mu} \cdot 100 \quad (11)$$

where ρ [$kg \cdot m^{-3}$] is the fluid density, g [$m \cdot s^{-2}$] is gravitational acceleration, and μ [$Pa \cdot s$] is the dynamic viscosity. Hydraulic conductivity is important as it dictates the velocity field. For single phase flow, heterogeneity has largely been modelled using porosity dependent relationships, however in multiphase flow modelling scaling factors are commonly used instead.

Porosity-based Permeability Correlation

There are many correlation used to generate a relationship between porosity, ϕ , and permeability, k . Some common examples include the Kozeny-Carman model[15] and the percolation theory, which are able to capture overall trends in heterogeneity, but lack the detail required by this study.

Fractal geometry is the most relevant model used to calculate hydraulic conductivity from previous studies. In this study, a combination [16] of three fractal geometry models are implemented as separate terms to respond to a range of porosities.

$$\mathbf{K} = S \cdot (6.2\phi_{mob} + 1493\phi_{mob}^2 + 58(10 \cdot \phi_{mob})^n) \quad (12)$$

Where, S is a dimensionless scaling factor applied to each voxel to match the average hydraulic conductivity of Ketton limestone, $K_{\text{mean}} = 1.888 \times 10^{-3} [\text{cm} \cdot \text{s}^{-1}]$ [17]. The exponent n controls how strongly the hydraulic conductivity responds to high-porosity voxels and is adjusted to tune the heterogeneity representation.

Scaling Factor

Scaling factors are a well established proxy for modelling heterogeneity in multiphase flow. They are generated by using X-Ray Computed Tomography (XCT) scans to map fluid saturation at different flow rates. The 3 dimensional saturation data is resampled into voxels at multiple resolutions (7–70 mm) to capture heterogeneity at multiple scales [7].

The scaling factor field is put into seven equal-sized bins between 0.7 and 1.3 to increase the heterogeneity. While the scaling factor is a promising model for the distribution of heterogeneity, it gives little information about the variance in the hydraulic conductivity. Because of this, the distribution is translated to hydraulic conductivity using a power law equation, which ensures a set experimental mean, while the variance can be used as a fitting parameter.

$$\mathbf{K} = K_{\text{mean}} \left(\frac{\mathbf{S}_f}{S_{f,\text{mean}}} \right)^B \quad (13)$$

Where, S_f is the scaling factor field, $S_{f,\text{mean}}$ is the calculated geometric mean 0.939. B is a parameter that can be varied to increase the variance of hydraulic conductivity. An exponent $B = 1$ would mean a completely linear translation with no stretching, while B larger than 1 would increase the distribution along the x-axis as seen in Figure 2. Furthermore, as the mean hydraulic conductivity is close to zero, as B is increased, the histogram tails to the right.

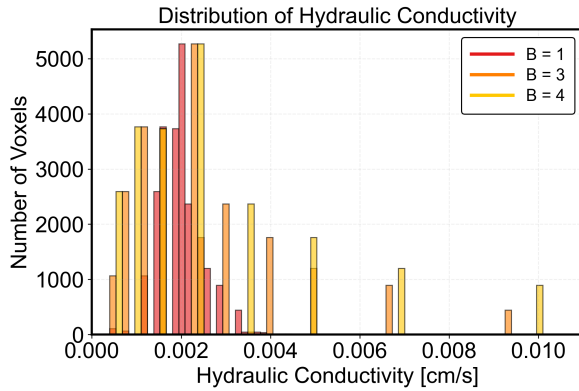


Figure 2: Histogram of the values of hydraulic conductivity derived from scaling factor correlations, by voxel at $B = 1, 3, 4$

4.5 Residence Time Distribution

Breakthrough curves in this study are calculated as the mean concentration of the final voxel-slice values for both the PET and model concentration arrays. Plotting the outlet concentration allowed a visual analysis of the spatial distributions.

For a quantitative analysis, the mean residence times, τ , of the models are calculated and compared to the experimental, τ_{exp} .

$$\tau = \frac{\int_0^\infty t C(t) dt}{\int_0^\infty C(t) dt} \quad (14)$$

4.6 Grid Sensitivity Analysis

Grid size can affect the accuracy of the model, as larger grid sizes lose detail. Grid sizes ranging from 0.1–5 [cm] are tested on a homogeneous synthetic system. The mean residence times, τ are then compared to the theoretical mean residence time of an ideal plug flow system, τ_{theo} .

$$\tau_{\text{theo}} = \frac{L \phi_{\text{mob}}}{u} \quad (15)$$

The mesh is considered "good" when the τ of the model is within 5[%] of the τ_{theo} . The second factor to consider when selecting a mesh size is the computational time required by the model. When voxel size fell below 1.2 [mm], computation time rose exponentially above five minutes. As the introduction of heterogeneity and mass transfer will further increase the computational time, five minutes is taken as the threshold for homogeneous systems. A voxel size of 1.66 x 1.66 x 1.66 [mm] is used throughout the study.

4.7 Moments

Other analytical methods include calculating spatial moments to describe tracer spread within the cylinder. This provides a measure of inner plume spread not just from the outlet, and displaying subtle, otherwise undetected nuances due to heterogeneity.

The zeroth moment is defined as:

$$M_0 = \int C dz \quad (16)$$

which acts as a confirmation for conservation of mass, and highlights when the total volume of tracer has reached the end of the sample length. M_0 is normalised by the theoretical tracer volume,

$$V_{\text{theo}} = u \cdot A \cdot t_{\text{inj}} \quad (17)$$

Where and t_{inj} represents the tracer injection time [s]. The first moment is defined as:

$$M_1 = \int z \cdot C dz \quad (18)$$

and describes where the center of mass lies within the sample through the model run. The first moment is normalised by the zeroth moment and the length of sample. Furthermore, the slope of this curve represents the interstitial velocity, v . The second moment is defined as:

$$M_2 = \int (z - \bar{z})^2 \cdot C dz \quad (19)$$

with the centroid $\bar{z}$ being:

$$\bar{z} = \frac{M_1}{M_0} \quad (20)$$

This metric describes how much the tracer plume deviates from its average flowpath, essentially the flow's variance [18]. The slope of this metric is equal to twice the longitudinal dispersion coefficient, D_L Eq.(5). The first moments is normalised by the zeroth moment, M_0 , before all three moments are normalised by the cylinder length. All moments are plotted dimensionless and the time domains are normalised by their mean residence times.

5 Results

5.1 Synthetic Model

This analysis is carried out to confirm the model and its corresponding packages used the input parameters correctly. It also served as a numerical insight into the effects of varying key model parameters.

Flow

The effect of velocity on the spread of tracer is tested via the Péclet number, The Péclet number expresses how strongly advection dominates over diffusion. In the ranges 80-600 ex-

plored in this study, flow is mainly advection dominated.

$$Pe = v \cdot \frac{d_p}{D} \quad (21)$$

where d_p us the grain size 600 $[\mu\text{m}]$ [10]

Figure 3a shows a shift in the mean residence time with a lower Péclet number and a wider temporal spread, this is because the breakthrough itself is slower. However, there is increased variance in the system with higher flow rates. The longitudinal dispersion or variance comes mostly from the increased advection at higher flowrates. As longitudinal dispersivity is constant the gradients of the second moment are all equal.

Finally, the Péclet number analysis confirms the model's physical validity. Table 1 shows that the mean residence times closely align with those predicted for ideal plug flow, while the moment analysis demonstrates that the observed gradients agree with analytical predictions and that mass is conserved throughout the simulations.

Table 1: Comparison of theoretical, τ_{theo} , and model mean residence times, τ , for Péclet numbers 80, 200, 400, 600. All models run using fractal geometry correlations ($n=10$) with $\alpha_l = 0.06$ [cm] and ($k = 0$ [s^{-1}])

Pe	τ_{theo} [min]	τ [min]
80	12.50	12.70
200	5.00	5.37
400	2.50	2.70
600	1.67	1.90

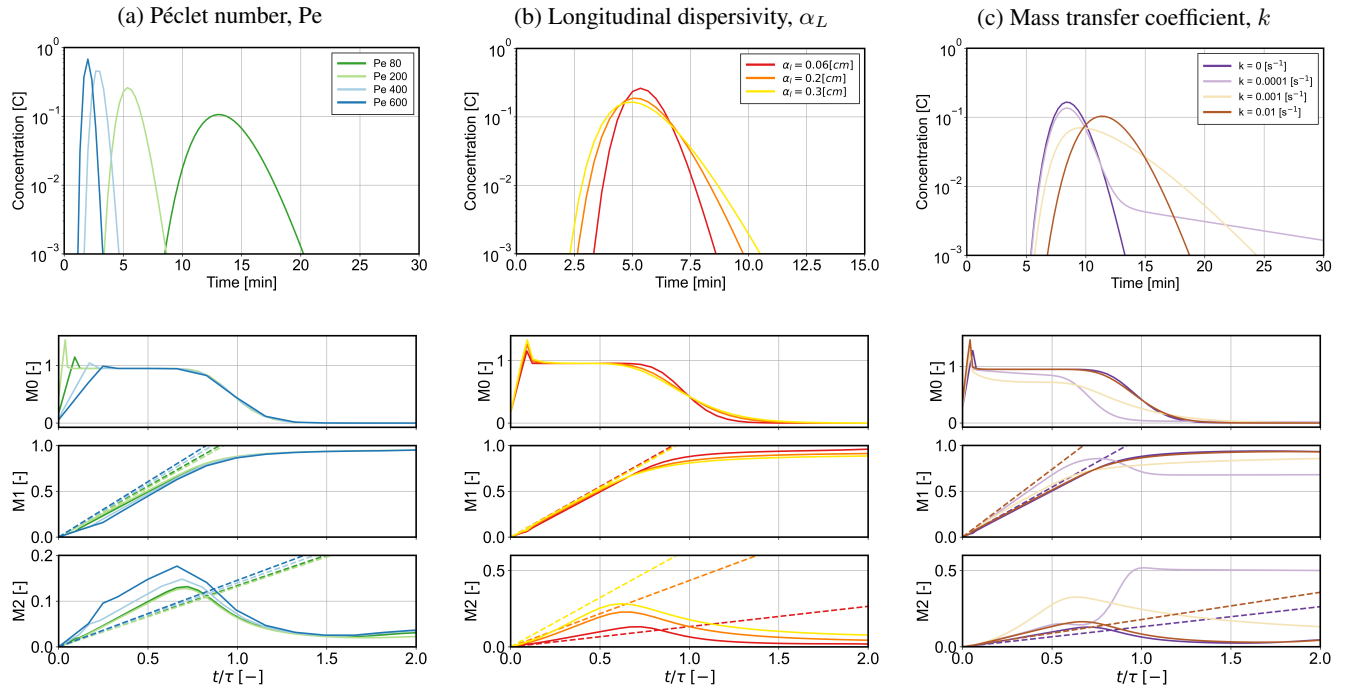


Figure 3: The figure shows the breakthrough curves (top) and moments (lower) for the following synthetic porosity cases: (a) $Pe = [80, 200, 400, 600]$, $\alpha_l = 0.06$ [cm], $k = 0.0$ [s^{-1}] (b) $Pe = 200$ $\alpha_l = [0.06, 0.2, 0.3]$ [cm], $k = 0.0$ [1/s] (c) $Pe = 200$, $\alpha_l = 0.06$ [cm], $k = [0.0, 0.0001, 0.001, 0.01]$ [s^{-1}]. All models were run with the fractal geometry model ($n=10$) an initial concentration $C_i = 1$ [-] and normalised according to Section 4.7. The dashed lines are the analytical gradients in Section 4.7.

Transport

Longitudinal dispersivity, α_L , is shown to have a little effect on the mean residence time, but a much larger effect on temporal spread, as seen in Figure 3b. This result is consistent as the theoretical mean residence time, Eq. (14), is independent of α_L . This longitudinal dispersivity analysis displayed a higher value may increase temporal spread, but only a value close to the grain size is appropriate [7].

The other input transport parameter is the mass transfer coefficient, $k[s^{-1}]$, the rate of material transfer from the immobile to mobile domain. In literature [8], it has been shown for this type of limestone, $k = 0.01 - 0.0001 [s^{-1}]$ are realistic values for this parameter. From Figure 3c, it is evident that this parameter has the largest effect on breakthrough time.

At $k = 0[s^{-1}]$, the system acts as a single porosity model according to the mobile domain with minimal spread, a mostly uniform curve. The smallest nonzero value, $k = 0.0001[s^{-1}]$, acts in a similar way up until around 11.5 [minutes], where there is a slow linear decline from then on. At such a low rate of mass transfer to mobile pores, it would appear that there is only mobile transport, until there is almost a second breakthrough from the immobile domain. It is only after this time that material from the immobile domain is able to trickle out the end of the cylinder. After increasing the rate of transfer, the system approaches a uniform curve, similar to that of a single porosity system. Carrying out this sensitivity analysis also confirmed the software is correctly inputting parameters and adhering to mass transfer laws present in the corresponding packages.

5.2 Experimental Data Application

Input porosity maps are introduced from experimental X-CT datasets. The corresponding mobile and immobile porosity fields are generated according to equation (10). The distribution obtained gave a mean total porosity, $\phi_t = 0.234$ and a mean mobile porosity, $\phi_{mob} = 0.163$, from a constant immobile porosity, $\phi_{im} = 0.085$. Concentration datasets from PET imaging are extracted to produce the analytical metrics

such as breakthrough curves and spatial moments necessary for model validation. This study used an inlet flowrate or superficial velocity, u , of $4[ml \cdot min^{-1}]$ corresponding to a Péclet number $Pe = 125.06$.

When plotting the breakthrough curves and moments in Figure 5, several key features are observed. The breakthrough curve exhibits a long tail on the right, the first moment shows a dip following the initial mean residence time, and the second moment displays two distinct gradients. These trends correspond to the shapes of mass transfer coefficients in the range $0.0001-0.001 [s^{-1}]$, as shown in 5c. In this regime, tracer exits the macropores while a significant portion remains trapped in the micropores. Once the initial tracer volume has left the system, the tracer slowly diffusing from the micropores causes the center of mass to shift back toward the domain center, explaining the dip in the first moment and the double gradient observed in the second moment.

Finally, the first three data points of the model are strongly influenced by entry effects. Although the model assumes that pure tracer is injected at time zero with no initial mixing, this is not physically achievable in practice. As a result, the first 3 points of the breakthrough curve present extended front tails. In the first moment, not all tracer is effectively injected until the third data point, and in the second moment, the variance shows a small initial spike.

Fractal Model Distribution

Varying the third exponent term, n , between 6 and 10 had a minor effect on the breakthrough curve, as shown in Figure 5a. Immediately, there is clearly a mismatch between the model curves and the experimental data. Despite this, increasing the exponent brings the model closer to the slope of the experimental tail, suggesting an even higher value would be the ideal numerical fit, however, this is not within a realistic range according to [19].

The moments echo this, with $n = 10$ best fitting with the experimental data. The first moment in particular shows an excellent fitting for the interstitial velocity, v , through the slope, however transport is not being modelled in the same

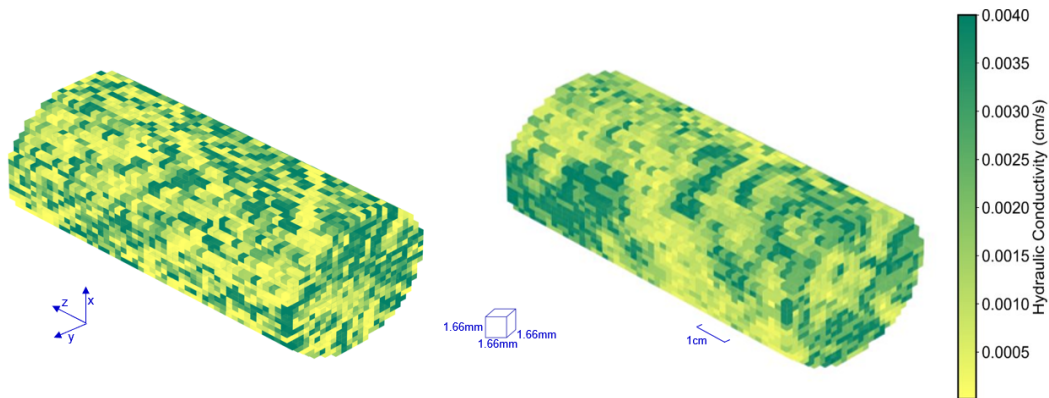


Figure 4: Hydraulic conductivity map. Left hand side: generated from the fractal geometry model with $n=10$. Right hand side: generated from scaling factor hydraulic conductivity map with $B = 3$. Both maps have the same dimensions (length: 10 cm, diameter: 5 cm, voxel size 1.66 x 1.66 x 1.66 mm)

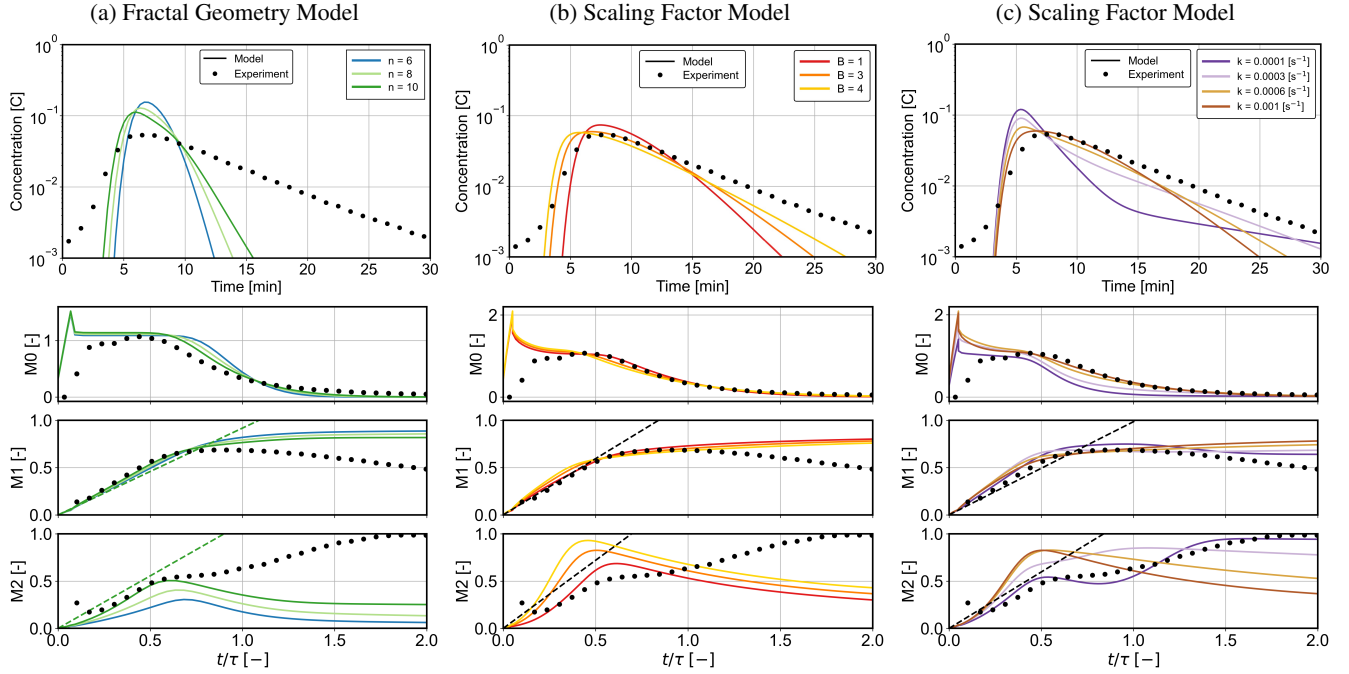


Figure 5: The figure shows the breakthrough curves (top) and moments (lower) for the following cases: (a) fractal geometry model at $n = [6, 8, 10]$, $k = 0.001 [s^{-1}]$ (b) scaling factor model $B = [1, 3, 4]$, $k = 0.001 [s^{-1}]$ (c) scaling factor model $B = 3$, $k = [0.0001, 0.0003, 0.0006, 0.001] [s^{-1}]$. All models were run with $\alpha_l = 0.06 [cm]$, initial concentration $C_i = 1 [-]$ at flow rate of $4 [ml \cdot min^{-1}]$; all model data is normalised according to Section 4.7. In the breakthrough curve, the experimental data is shifted so that $\tau = \tau_{theo}$. The moments of the experiment are normalised by its own τ . The dashed line in M1 and M2 graphs is the analytical gradient calculated in Section 4.7.

Table 2: Table shows the following values derived from the breakthrough curve mean residence times (τ [min]), time at maximum concentration ($t|_{C=C_{max}}$ [min]), maximum concentration ($C_{max} [-]$), and the least squares error $m = (C_{experiment} - C_{model})^2$. Columns 1–5, 6–10, and 11–15 correspond to Figures 5a, 5b, and 5c, respectively.

n	τ [min]	$t _{C=C_{max}}$ [min]	C_{max} [-]	m	B	τ [min]	$t _{C=C_{max}}$ [min]	C_{max} [-]	m	k	τ [min]	$t _{C=C_{max}}$ [min]	C_{max} [-]	m
Experimental Data														
-	11.8	7.5	5.35e-2	0	-	11.8	7.5	5.35e-2	0	-	11.8	7.5	5.35e-2	0
Model Data														
6	7.34	6.68	1.55e-1	2.27e-2	1	9.77	7.38	5.35 e-1	1.68e-3	1e-4	8.1	5.37	1.20 e-2	1.62e-2
8	7.34	6.34	1.28e-1	1.39e-2	3	9.94	6.71	5.95 e-2	9.53e-4	3e-4	9.5	5.37	9.00 e-2	6.97e-3
10	7.34	6.01	1.12e-1	9.18e-3	4	9.86	5.70	5.77 e-2	2.03e-3	6e-4	9.8	5.70	6.80 e-2	2.41e-3
										1e-3	9.9	6.70	5.95 e-2	9.53e-4

way since the model does not slope downwards like the experiment. Overall this analysis shows that modifying the fractal geometry model has a small effect on transport, and continues to demonstrate a gap between experimental data and modelled values.

Scaling Factor Distribution

The B coefficient, as defined in equation (13), is tested in the range 1–4 with a mass transfer coefficient of $0.001 [s^{-1}]$. As shown in Figure 5b, increasing B extends the tail of the breakthrough curve to the right. However, this also shifts the mean residence time to the left, moving it further from the theoretical value. Compared to the fractal geometry model, the breakthrough curves with the B -scaling approach more closely match the experimental shapes. While $B = 4$ best re-

produced the tail, $B = 3$ exhibits the least deviation from the experiment according to the least squares error. When considering the second moment the model does not capture changes in the gradients characteristic of the micro pores in the experimental data.

Secondary mass transfer coefficient fitting

To better fit the second moment mass transfer coefficient, k , is explored in the range 0.0001 – $0.001 [s^{-1}]$, based on the initial sensitivity analysis, with a B value of 3. As shown in Figure 5c the lowest value, $k = 0.0001 [s^{-1}]$, produces a high early peak and a sudden drop in concentration, but provides the best fit for the first and second moments. A slightly higher value, $k = 0.0003 [s^{-1}]$, reduces early deviation and accurately reproduces the slope of the breakthrough

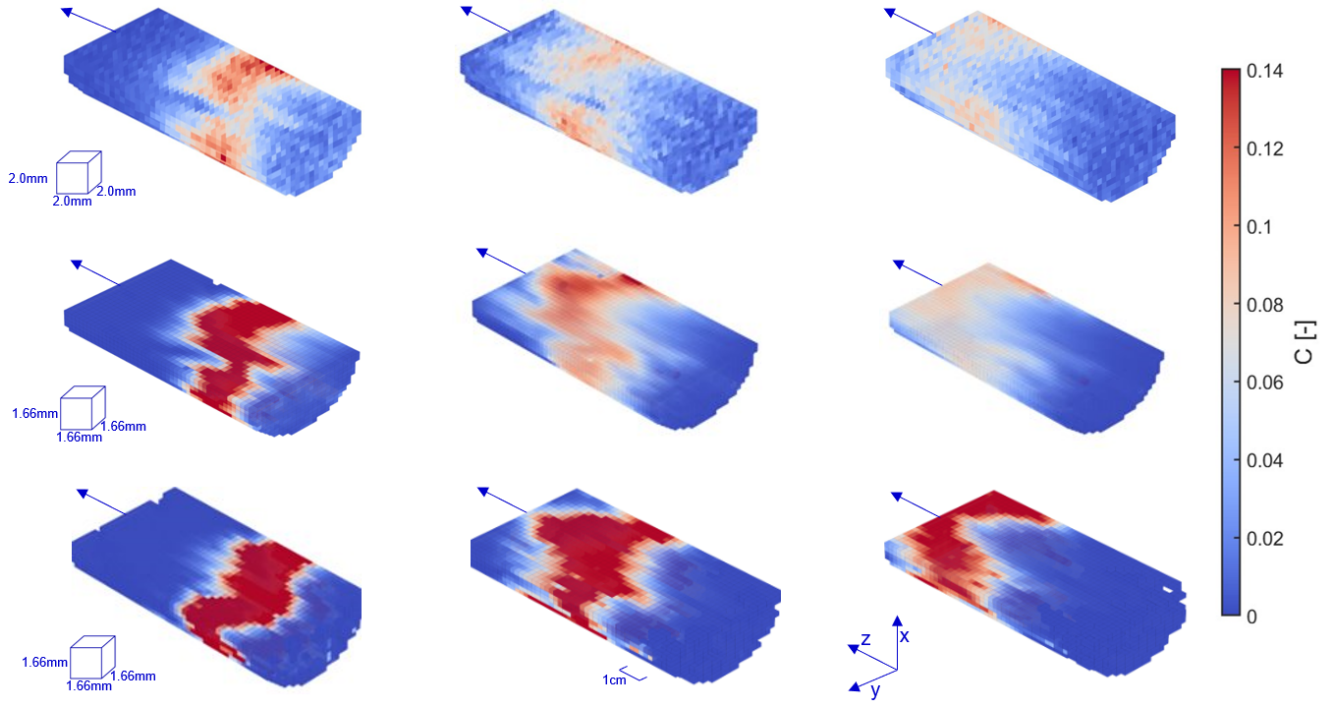


Figure 6: The lower hemisphere of the 3D concentration plots are shown at times 2.5, 4.5, 6.5 [min] (left to right) All plots use a flowrate of $4 \text{ [ml} \cdot \text{min}^{-1}]$ and are normalised to $[ml_{\text{tracer}}/ml_{\text{fluid}}]$. First row is experimental data from PET imaging [11] Frames 4, 6, 8 respectively. Row 2 is the model using scaling factor, $B=3$. The third row is uses fractal geometry with $n=10$. The models use $k = 0.001 \text{ [s}^{-1}]$, $\alpha_l = 0.06 \text{ [cm]}$ and a voxel size of $1.66 \times 1.66 \times 1.66 \text{ [mm]}$.

curve's right tail, though it still peaks early. The highest value, $k = 0.001 \text{ [s}^{-1}]$, best aligns with the experimental breakthrough curve up to 15 [minutes] and yields the smallest error.

5.3 3 Dimensional Concentration Maps

3 dimensional concentration maps are generated to visually display the "best" (least error) model in both fractal geometry and scaling factor model. The PET data is normalised by the total inlet concentration to produce dimensionless values.

The simulated concentration plots are initially higher than the experiment, possibly due to the model's perfect injection and lack of entry effects. As shown in Figure 6, the scaling factor model (row 2) closely matches the experimental shape tailing streaks and the overall dispersion pattern (row 1). In contrast, the fractal geometry model (row 3) follows the general shape but fails to reproduce the characteristic longitudinal dispersion and high-concentration streaks seen in the experimental data.

6 Discussions

6.1 Structure Heterogeneity

When considering heterogeneity, the porosity seems like the obvious correlation, and is traditionally used in single phase flow [6]. However porosity maps can only give us an indi-

cation of the fraction of space available in each voxel and not the pore interconnectedness. If pore space is high, yet there is no pathway to, it this space cannot contribute towards flow. In contrast, scaling factor data is derived from saturation maps. Saturation is controlled by the rate and extent that the fluid can reach a certain area, and therefore, is a better indication of pore connectedness. In Figure 4, the map of hydraulic conductivity from scaling factor has larger channels, a better description of where fluid is able to travel, than the more disordered fractal geometry derived hydraulic conductivity map. Our data validates this theory, as the scaling factor model not only achieved the least error compared to the breakthrough curve, but also visibly matches the flow patterns in Figure 6.

6.2 Dual Porosity

Accounting for micropores is one of the main features that sets this study apart from traditional porous modelling. Without mass transfer, the breakthrough curve approaches a near-symmetric, parabolic shape. The interaction with micropores is what causes the characteristic tailing to the right of the breakthrough curve seen in experimental data, as well as the secondary slope present in the second moment. As the bulk of the concentration exits the system, breakthrough becomes limited by the mass transfer between mobile and immobile domains. Qualitatively, we can confirm that mass transfer coefficients in the range $0.0001\text{-}0.001 \text{ [s}^{-1}]$ mimic these dispersion signatures, reinforcing the need for considering dual porosity when modelling solute transport in porous media.

6.3 Model Limitations

Although this model generated numerous numerical and visual metrics comparable to 3 dimensional experimental datasets, it is still limited in various areas. Unfortunately, this academic study was mainly limited by computational time of the heterogenous model, upwards of 20 [minutes] per run. Consequently, a comprehensive parameter optimisation was not in the scope of this project and instead an informed, realistic range was explored. This also proved detrimental to the

grid sensitivity analysis, as although there was an exponential increase in computational time below a certain size, a finer grid size may have improved precision regarding permeability and concentration fields.

Looking at the dual porosity model, in this study like many others, fluid in the microporous region is considered stagnant, meaning microporous interconnectivity is neglected. This then limits permeability heterogeneity, as slower microporous flow is not taken into account, which could reduce the dispersion accuracy captured in 3D plotting.

7 Conclusion

This study was conducted to model solute transport in a microporous limestone via MODFLOW-2005 under a flowrate of $4 [ml \cdot min^{-1}]$. The model incorporated both synthetic and XCT-derived porosity fields. Tracer pulse experiments were simulated using the 3 dimensional models. Most studies validate result purely using breakthrough curves (1 dimensional), for a more detailed approach, we generated 3 dimensional concentration fields from PET-imaged experimental data, to compare our models. This allowed for quantitative and qualitative comparison, addressing a gap in result validation.

Second a dual domain system is implemented via separate mobile and immobile porosity fields with intragranular mass transfer, k , between them. This mass transfer allowed the model to capture key concentration signatures, such as delayed tracer release, long breakthrough curve tailing and a secondary slope within the second spatial moment. By implementing the interactions between different sized pores 3 dimensional concentration fields accurately mapped dispersion, and tailing.

Most importantly, 3 dimensional permeability fields are generated from both fractal geometry and scaling factor correlations. While the fractal geometry model could capture general trends like the centre of mass, it failed in predicting temporal spread or reproduce PET imaged plume streaks and dispersion. In comparison, the scaling factor correlation adapted from multiphase flow studies, proved to be more flexible and accurate at capturing experimental dispersion and dual-porosity characteristics, with $B = 3$ and $k = 0.0001 - 0.001[s^{-1}]$ closely aligning with experimental τ , breakthrough curve tailing and 3 dimensional PET visual dispersion. This demonstrated that scaling factor heterogeneity gives a more realistic pore connectivity model in Ketton limestone, and produced the closest alignment across validation metrics ($m < 10^{-3}$).

Overall, this project highlighted that accurate 3 dimensional modelling of solute transport through Ketton limestone requires a distinguished dual porosity system and that scaling factor based heterogeneity to deliver a realistic predictions. This addressed the research gap between 1 and 3 dimensional modelling, and sets a basis for 3 dimensional validation.

This study indicates that in-depth optimisation of the model parameters is necessary. Future work should include a dedicated Python-based optimisation workflow capable of systematically exploring parameter space, such as permeability scaling, dispersivity, and mass, transfer coefficients, to quantitatively align the simulated and experimental spatial moments as well as the full 3 dimensional concentration fields. Such a framework would reduce manual tuning, and provide a more accurate representation of concentration field. Another future recommendation explored in previous studies [20], is the concept of a dual permeability approach. This system would theoretically generate two separate permeability fields corresponding to macroporosity and microporosity, then later overlaid in the model input[20]. This would consider how microporosity plays a role in the conductivity field, possibly improving interconnectivity.

Acknowledgement

We would like to thank Takeshi Kurotori, who helped us with this project.

References

- [1] A. Laoufi et al. "Integrated Simulation of Groundwater Flow and Nitrate Transport in an Alluvial Aquifer Using MODFLOW and MT3D: Insights into Pollution Dynamics and Management Strategies". In: *Sustainability* 16 (2024), p. 10777. DOI: 10.3390/su162310777. URL: <https://doi.org/10.3390/su162310777>.
- [2] P. Li et al. "Sources and Consequences of Groundwater Contamination". In: *Archives of Environmental Contamination and Toxicology* 80.1 (2021), pp. 1–10. DOI: 10.1007/s00244-020-00805-z.

- [3] A. Khanal and M. I. Khan. “Impact of Relative Permeability Hysteresis and Capillary Pressure on Trapping Mechanisms During CO₂ Sequestration in Saline Aquifers”. In: *Proceedings of the Carbon Capture, Utilization, and Storage Conference (CCUS)*. CUS: 4012176. Carbon Capture, Utilization, and Storage Conference. Houston, TX, USA, Mar. 2024. DOI: 10.15530/ccus-2024-4012176.
- [4] J. Ubillus et al. “Impact of small-scale heterogeneity on field-scale CO₂ migration and trapping”. In: *Proceedings of the 17th Greenhouse Gas Control Technologies Conference (GHGT-17)*. Presented 20–24 October 2024. Nov. 2024. DOI: 10.2139/ssrn.5034892. URL: <https://ssrn.com/abstract=5034892>.
- [5] S. J. Jackson and S. Krevor. “Small-Scale Capillary Heterogeneity Linked to Rapid Plume Migration During CO₂ Storage”. In: *Geophysical Research Letters* 47.18 (2020), e2020GL088616. DOI: 10.1029/2020GL088616.
- [6] M. Krause, J.-C.-C. Perrin, and S. M. Benson. “Modeling permeability distributions in a sandstone core for history matching coreflood experiments”. In: *SPE Journal* 16.4 (2011), pp. 768–777.
- [7] T. Kurotori and R. Pini. “A general capillary equilibrium model to describe drainage experiments in heterogeneous laboratory rock cores”. In: *Advances in Water Resources* 152 (2021), p. 103938. DOI: 10.1016/j.advwatres.2021.103938.
- [8] T. Kurotori et al. “Measuring, imaging and modelling solute transport in a microporous limestone”. In: *Chemical Engineering Science* 196 (2019), pp. 366–383. DOI: 10.1016/j.ces.2018.11.001.
- [9] W. Song et al. “Pore scale modeling of fluid transport in complex reservoirs: Multi-scale digital rock construction, flow experiments and simulation methods”. In: *Capillarity* 11.3 (2024), pp. 81–88. DOI: 10.46690/capi.2024.06.03. URL: <https://www.sciopen.com/article/10.46690/capi.2024.06.03>.
- [10] T. Kurotori et al. “Direct observations of solute dispersion in rocks with distinct degree of sub-micron porosity”. In: *Water Resources Research* 61 (2025). Received 12 AUG 2024, Accepted 29 JAN 2025, e2024WR038625. DOI: 10.1029/2024WR038625.
- [11] S. D. Repository. *Ketton Limestone PET and XCT Imaging Dataset*. Stanford Digital Repository. Available at <https://purl.stanford.edu/ts931rx0407>. 2025.
- [12] U. M. D. Team. *MODFLOW 6 Documentation*. <https://modflow6.readthedocs.io/>. Accessed: 2025-12-07.
- [13] A. Honari et al. “Enhanced gas recovery with CO₂ sequestration: the effect of medium heterogeneity on the dispersion of supercritical CO₂–CH₄”. In: *International Journal of Greenhouse Gas Control* 39 (2015), pp. 39–50.
- [14] M. Boon, B. Bijeljic, and S. Krevor. “Observations of the impact of rock heterogeneity on solute spreading and mixing”. In: *Water Resources Research* 53.6 (2017), pp. 4624–4642. DOI: 10.1002/2016wr019912.
- [15] B. H. Lodhia and S. R. Clark. “Computation of vertical fluid mobility of CO₂, methane, hydrogen and hydrocarbons through sandstones and carbonates”. In: *Scientific Reports* 12 (2022), p. 10216.
- [16] M. Krause, J.-C. Perrin, and S. M. Benson. “Modeling Permeability Distributions in a Sandstone Core for History Matching Coreflood Experiments”. In: *SPE International Conference on CO₂ Capture, Storage, and Utilization*. SPE 126340. Society of Petroleum Engineers. San Diego, California, USA, Nov. 2009.
- [17] R. Pini. *Fourth Year Research Project Specification for “Modelling of Three-Dimensional Solute Transport in Disordered Porous Media”*. Tech. rep. Supervisors: Prof Ronny Pini and Dr Takeshi Kurotori. Students: Erin Richards, Eszter Gyuris. Imperial College London, June 2025. URL: <mailto:r.pini@imperial.ac.uk>.
- [18] G. Gundersen. *Understanding Moments*. <https://gregorygundersen.com/blog/2020/04/11/moments/>. Accessed: 2025-12-08. 2020.
- [19] H. Pape, C. Clauser, and J. Iffland. “Permeability prediction based on fractal pore-space geometry”. In: *Geophysics* 64.5 (1999), pp. 1447–1460.
- [20] A. Balogun et al. “Verification and Proper Use of Water/Oil Transfer Function for Dual-Porosity and Dual-Permeability Reservoirs”. In: *SPE Middle East Oil and Gas Show and Conference*. Society of Petroleum Engineers. Manama, Bahrain, Mar. 2007.

Mathematical Modeling of Spray-Based Microalgae Culture Systems in Column Photobioreactors

Ammar Azmi, ChengXuan Cheong

Department of Chemical Engineering, Imperial College London, U.K

Abstract

Microalgae cultivation in photobioreactors (PBRs) is fundamentally governed by light attenuation through the culture medium, resulting in non-uniform irradiance distribution and substantial losses in biomass productivity compared to theoretical maximum rates. Conventional designs such as raceway ponds and large-scale closed PBRs struggle to balance competing constraints of photoinhibition in highly illuminated and severely darker regions in the culture. This work presents a comprehensive mathematical model of a novel spray-based photobioreactor (sb-PBR) that integrates spray dynamics, light distribution, and photosynthetic growth kinetics into an integrated simulative framework. The model couples self-similar turbulent jet spray dynamics with radiative transfer equations incorporating Mie theory scattering effects and a physiologically-resolved photosynthesis kinetics model accounting for photoacclimation and non-photochemical quenching responses. Spray dynamics are modeled using a two-stage formulation with droplet concentration determined from mass conservation principles. The light propagation is solved using the two-flux approximation for a defined discretized vertical layer. Lastly, the growth kinetics is described using a system of equations that captures dynamic states of Photosystem II and photoregulation mechanisms of microalgae. The system of coupled nonlinear ordinary and partial differential equations is implemented in Python with numerical integration via appropriate solvers, the MAGNUS package. Results demonstrate that spray-column configuration achieves improved biomass growth to conventional systems through short optical path enhanced surface-to-volume ratios inherent to droplet-based cultivation. The integrated model serves as a predictive design tool for systematic exploration of operational parameters including dilution rate, nozzle geometry, column height, and incident irradiance. This allows for identification of optimal cultivation conditions. This modeling framework provides quantitative guidance for rational design of high-performing closed PBRs for commercial-scale microalgae production.

Keywords: MCS, PBR, sb-PBR, light attenuation, photo-acclimation, NPQ, MAGNUS

1. Introduction

Microalgae are unicellular photosynthetic organisms characterised by exceptionally high photosynthetic efficiency, converting solar energy into chemical energy through atmospheric CO₂ fixation via photosynthesis. Under photoautotrophic conditions, microalgae utilise light as the primary energy source and carbon dioxide as the sole carbon source (Razzak et al., 2023). Microalgae group encompasses both prokaryotic cyanobacteria and eukaryotic taxa including Chlorophyta, Rhodophyta, and Bacillariophyta. Microalgae are valued across food, feed, and cosmetics industries owing to their capacity to synthesise a diverse array of biologically active compounds, ranging from polysaccharides, proteins, and lipids to photosynthetic pigments as well as bioactive secondary metabolites (Geng et al., 2025). As renewable biofuel feedstocks, microalgae present a compelling alternative to fossil fuels in terms of economic viability, renewability, and environmental sustainability (Khan et al., 2018).

PBR performance is fundamentally governed by three

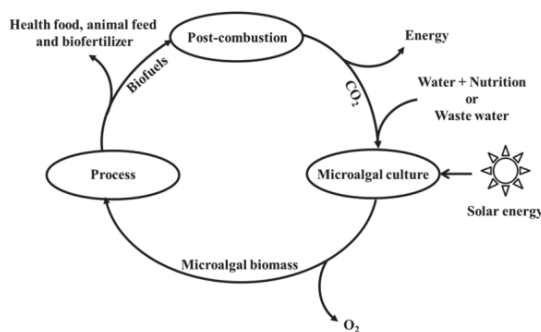


FIGURE 1: An integration of microalgae cultivation with CO₂ utilization and their applications as biofuels (Klinthong et al., 2015)

coupled phenomena : gas-liquid mass transfer, mixing dynamics, and light availability within the culture. Microalgae consume dissolved CO₂ for carbon fixation whilst releasing dissolved O₂ as a metabolic by-product. Continuous, efficient exchange of these gases across the gas-liquid interface is essential to maintain adequate substrate

supply and prevent inhibitory oxygen accumulation. Mixing dynamics directly determine the residence time of cells within distinct light regimes. Classical PBR theory demonstrates that intermittent exposure to alternating light and dark zones enhances overall growth compared to constant low-light conditions (Fathurrahman et al., 2013). However, excessively rapid light-dark cycling results in insufficient cellular repair and growth, whilst slow cycling permits photodamage and oxygen depletion.

Raceway ponds currently represent the predominant large-scale microalgae cultivation system being deployed and configured as closed-loop channels operating in open-air conditions. Although economically attractive, raceway ponds exhibit inherent susceptibility to environmental fluctuations, evaporative water loss, and biological contamination due to their open configuration. Furthermore, raceway ponds are characterised by low biomass productivity, primarily attributable to severe light attenuation challenges throughout the dense culture medium (Benner et al., 2022). Conversely, closed PBRs offer superior biomass productivity, enhanced contamination control, and high reproducibility. These advantages render closed PBR systems preferable for commercial-scale operations and production of high-value specialty compounds (Changuia et al., 2021).

Mathematical model plays a pivotal role in the design of PBRs due to the multiphysics of microalgal culture systems that couples biological growth with light-transfer, hydrodynamics and mass transport. Lab-scale observations cannot capture the complex feedback systems between biomass accumulation and light attenuation along with the rapid fluctuating light environments experienced by cells due to mixing (Bernard et al., 2016). Model-based optimization techniques are critical to estimate the maximum biomass production, enabling efficient operation and control for industrial-scale microalgae culture systems.

This research aims to develop a comprehensive mathematical model of a vertical spray-column PBR that integrates the coupled dynamics of light distribution, spray dynamics, and microalgal growth kinetics. Spray columns represent a substantially underexplored reactor configuration with promising advantages, yet their design principles and optimisation strategies remain poorly characterised. Through implementation of this integrated model in Python utilising appropriate numerical solvers and optimisation algorithms, the relationships between spray hydrodynamics, droplet optical properties, light availability to cells, and resulting growth kinetics are defined. The model will serve as a predictive design tool enabling systematic exploration of the operational parameter space, including spray flowrate, nozzle geometry, column height, and incident irradiance. Consequently, this modelling work will enable identification of optimal cultivation conditions for microalgae in spray-circulated systems.

2. Background

The process of microalgae cultivation requires precise control of several key operational parameters such as light supply, temperature, pH, and dissolved oxygen and carbon dioxide concentrations, together with nutrient availability. In the context of light supply, photosynthetically active radiation (PAR) ranges from 400–700 nanometers, corresponding to the absorption maxima of primary photosynthetic pigments. Light attenuation within the culture medium occurs through scattering and absorption. This spatial heterogeneity in light distribution creates a non-uniform irradiance field throughout the PBR, translating to non-uniform growth kinetics. A critical strategy for mitigating these adverse effects is to model a system with minimal attenuation to increase light utilisation efficiency, thereby yielding a more uniform and predictable growth kinetics.

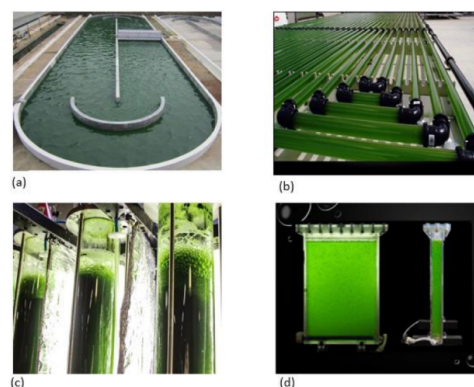


FIGURE 2: Common forms of photobioreactors. (a) an open pond system, (b) Tubular photobioreactor, (c) Bubble column photobioreactor and (d) Flat panel photobioreactor (Agbebi, 2019)

Closed PBRs enable excellent contamination control and precise process control over critical parameters, whilst demonstrating significantly higher biomass productivity and photosynthetic efficiencies compared to open systems (Benner et al., 2022). Several conventional closed PBR designs are well-established in current literatures : bubble columns, airlift reactors, flat-panel, and tubular configurations. Each design represents a distinct compromise among light utilisation, mass transfer efficiency, mixing performance, and scalability (Uyar et al., 2024). Bubble column reactors offer low operating costs and excellent scalability but suffer from non-uniform light distribution, whilst airlift reactors improve efficiency through buoyancy-driven circulation. Flat-panel reactors achieve good light utilisation but face scalability issues whilst tubular system designs provide modularity but present challenges in temperature control and biofilm formation.

An emerging alternative is the aerosol-based photobioreactors (ab-PBRs), wherein nutrient and water are supplied as fine aerosol mist rather than bulk liquid immersion, with microalgae cultivated in biofilm form (Schulze et al., 2014). This aerosol delivery approach substantially

increases the interfacial area for gas-liquid contact, thereby accelerating CO_2 transfer and oxygen removal, whilst permitting high cell density cultivation and environmental stress manipulation. However, substrate-based cultivation such as the one used in ab-PBRs presents significant light utilisation challenges. The spray-based photobioreactor (sb-PBR) proposed herein operates by spraying microalgae suspended in nutrient-rich medium as a fine mist through a vertical, cylindrical, transparent column with LEDs arranged circularly around the exterior and CO_2 gas introduced at the base. This emerging design exploits exceptional surface-to-volume ratios inherent to droplet-based cultivation, potentially enabling substantially improved light utilisation efficiency. However, current literature lacks comprehensive mathematical models describing the coupled interactions between spray hydrodynamics, light distribution and growth kinetics.

3. Methodology

3.1. System Configuration

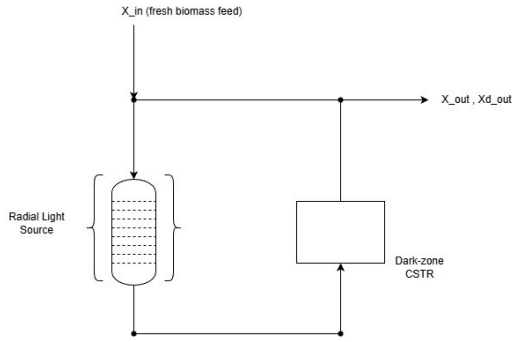


FIGURE 3: Proposed design configuration for sb-PBR system

The sb-PBR comprises of a vertical cylindrical column with circumferential LED illumination. Fresh culture medium is sprayed at the column top and descends as droplets through the illuminated region, whilst counter-current CO_2 -enriched gas is sparged from the base to facilitate gas-liquid mass transfer. The culture discharged from the column bottom is directed into a dark-zone CSTR for biomass sedimentation and harvesting, with clarified medium recycled to the column inlet. For this first approximation model, nutrient-replete conditions are applied while, concentration of CO_2 and temperature effects are ignored. A schematic of the proposed design can be referred to in figure 3. The operating parameters to be used in simulation is listed in table 3.1.

TABLE 1: Operating parameters for spray column photobioreactor

Parameter	Value
Column radius, r	0.25 m
Column height, H	2 m
Biomass Feed, X_{ref}	0.015 g/L
Recycle ratio, R	0.3
Incident light intensity, I_0	$100 \mu \text{ E m}^{-2} \text{ s}^{-1}$
Inlet liquid flowrate, Q_L	$5 \times 10^{-5} \text{ m}^3/\text{s}$
Gas flowrate, Q_G	$2.5 \times 10^{-3} \text{ m}^3/\text{s}$
Nozzle radius, r_0	$4 \times 10^{-6} \text{ m}$

3.2. Modelling Overview

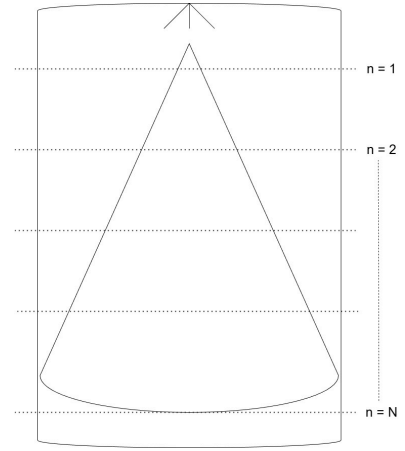


FIGURE 4: Spray discretization into N layers

The PBR spray column system is governed by a coupled set of nonlinear partial differential equations that describe light attenuation, photosynthetic kinetics, and biomass transport across vertically discretized layers. The discretization approach to be employed is visualized in figure 4. The system model consists of four integrated blocks that perform different functions as shown in figure 5. First, block P1 pre-computes spray geometry and droplet hydrodynamics to characterize the spatial distribution and hydrodynamics of the droplets in the PBR. Block 1 solves the two-flux radiative transfer equation to determine average irradiance as a function of biomass concentration and extinction coefficients from Mie theory. Then, Block 2 integrates the Nikolaou photosynthesis model [Nikolaou et al. (2016a)] to compute layer-specific local growth rates that account for dynamic photosynthetic apparatus states at specific time points. The final Block 3 couples continuous stirred-tank reactor (CSTR) transport dynamics with biomass growth and respiration kinetics to return a value for both active and dead biomass.

This system of equations is spatially heterogeneous where each layer simulates varying dynamic properties due to varying biomass and droplet number concentration. This set of equations is highly parametric, with model predictions sensitive to kinetic rate constants, extinction coefficients

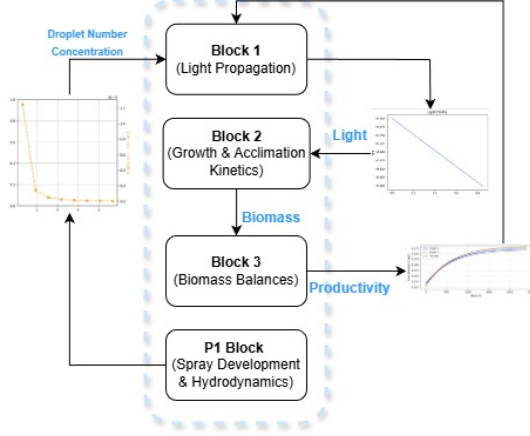


FIGURE 5: Coupling methodology of photosynthesis kinetics, hydrodynamics and light attenuation in spray column

and physiological parameters governing photosynthetic efficiency. As a first approximation model, the (Modeling, Analysis and desiGN of Uncertain Systems) MAGNUS package is used for structuring the mathematical framework. The MAGNUS library allows for seamless integration of the complex multi-block architecture and enables systematic exploration of parameter space to identify which model outputs are most sensitive to parameter variations. This forward uncertainty quantification capability is essential during early model development phase, as it reveals which parameters require refinement through literature review or future experimental campaigns, which physical processes dominate system behavior across different operating conditions and whether the proposed model structure can plausibly capture the expected effectiveness of sb-PBRs before investment in experimental validation efforts.

3.3. Spray Development Pattern

The spray column PBR directs a liquid spray downward against counter-current CO_2 gas flow, with droplets comprising medium and microalgal cells. The spray development is described by a two-stage geometric model grounded in jet and spray fluid mechanics. Experimental investigations using tomographic particle image velocimetry have revealed distinct spreading regimes as liquid jets atomize downstream from the nozzle (Buchmann et al., 2014). In the near-field region at approximately 12 nozzle diameters distance, the spray expands at a nearly constant angle, reflecting momentum-dominated behavior where the spray cone angle is determined by nozzle geometry and initial conditions. Beyond the transition height, the far-field region exhibits decreasing expansion rate as momentum dissipates through gas entrainment and inter-droplet collisions, with the spray losing coherence and the effective spreading angle contracting.

Transition point (empirical) :

$$z_{\text{transition}} = 12d_0 \quad (\text{self-similarity begins}) \quad (1)$$

Radius profile (two-stage) :

- Near-field ($z < z_{\text{transition}}$),

$$r(z) = r_0 + z \tan(\theta_{\text{initial}}) \quad (2)$$

- Far-field ($z \geq z_{\text{transition}}$),

$$r(z) = r_{\text{transition}} + (z - z_{\text{transition}}) \times 0.1 \quad (3)$$

Where :

$$r_{\text{transition}} = r_0 + z_{\text{transition}} \tan(\theta_{\text{initial}}) \quad (4)$$

Droplet number concentration :

$$n_0 = \frac{Q_{\text{droplet}}}{\pi r_0^2 v_{\text{net}}} \quad (5)$$

$$n(z) = n_0 \left(\frac{r_0}{r(z)} \right)^2 \quad (6)$$

The outputs from Block P1 return $n(z)$ values describing the droplet concentration profile in a discretized layer, derived from continuity conservation applied to the spray curtain. The inverse-square-law relationship as shown in equation 5 reflects geometric spreading. As the spray radius increases, droplets distribute over a larger cross-sectional area, reducing local concentration proportionally to area expansion. This concentration profile directly influences light attenuation, since extinction coefficients depend linearly on droplet density. The continuity-based approach is robust because it relies solely on mass conservation, requiring no empirical correlations or detailed velocity profile assumptions, ensuring physical consistency across different spray configurations and operating conditions.

3.4. Droplet Hydrodynamics

The vertical motion of droplets through the spray column is governed by the balance between gravitational and drag forces. In still air, droplets accelerate under gravity until viscous drag forces equilibrate the gravitational force, attaining terminal velocity. For a spherical droplet moving through a viscous fluid at low Reynolds number, Stokes' law provides an exact description of the drag force as $F_d = 6\pi\mu rv$, where μ denotes the dynamic viscosity of the surrounding medium and v is the relative velocity (Bird et al., 2007). At terminal velocity, this viscous drag balances the net gravitational force corrected for buoyancy. The validity of Stokes' law requires the low Reynolds number regime, characterized by $Re = \frac{\rho_{\text{air}} v_{\text{term}} d_d}{\mu_{\text{gas}}} \ll 1$, where inertial forces remain negligible compared to viscous forces. For the present system parameters, the calculated Reynolds number is approximately 0.067, confirming that droplet motion is governed purely by viscous drag with negligible contributions from form drag or turbulent separation effects.

In the spray column PBR, droplets encounter a continuous upward gas flow sparged from the column base. The

upward gas velocity is determined from continuity considerations as shown in equation [10](#), assuming uniform radial gas distribution across the column cross-section. The net downward droplet velocity is computed as the difference between the Stokes terminal velocity and the upward gas velocity, thereby capturing the opposing effects of gravity-driven settling and gas-phase drag. This approach is physically valid because the small droplet inertia ensures rapid local momentum equilibrium between the droplet and gas phases, rendering secondary effects such as Saffman lift force (significant only at $Re > 50$) or inter-droplet interactions negligible within the dilute spray regime. The complete formulation is shown in equations [7](#) - [12](#).

Applying force balance on droplets :

$$F_d = F_g \quad (7)$$

Substituting Stokes' drag law and gravitational force :

$$6\pi\mu r v_{\text{term}} = \frac{4}{3}\pi r^3(\rho_{\text{droplet}} - \rho_{\text{gas}})g \quad (8)$$

Stokes settling velocity :

$$v_{\text{term}} = \frac{g(\rho_{\text{water}} - \rho_{\text{air}})d_d^2}{18\mu_{\text{gas}}} \quad (9)$$

With upward gas flow :

$$u_{\text{gas}} = \frac{Q_{\text{gas}}}{\pi R^2} \quad (10)$$

Net downward velocity (constant for all layers) :

$$v_{\text{net}} = (v_{\text{term}} - u_{\text{gas}}) \quad (11)$$

Residence time per layer (constant across all layers) :

$$t_{\text{res}} = \frac{\Delta z}{v_{\text{net}}} \quad (12)$$

3.5. Light Distribution

3.5.1. Mie Theory

Light propagation through the spray column is significantly attenuated by absorption due to pigments within microalgae cells and by scattering from suspended water droplets containing the cells. The optical properties of microalgae cell are described by Mie theory using wavelength-resolved absorption and scattering cross sections for spherical particles with their size and complex refractive index. ([Sforza et al., 2020](#)). Mie theory from describes the scattering and absorption of electromagnetic radiation by spherical particles by solving Maxwell's equations for the interaction of a plane wave with a homogeneous sphere. The solution is expressed in terms of two fundamental parameters : the size parameter $x = \frac{2\pi r}{\lambda}$, which is the ratio of the particle circumference to the wavelength of incident light, and the relative refractive index $m = \frac{n_{\text{particle}}}{n_{\text{medium}}}$, which characterizes how the optical properties of the particle differ from those of the surrounding medium. The theory

yields expressions for the extinction Q_{ext} and scattering efficiency factor Q_{sca} , which represents the dimensionless ratio of the effective extinction or scattering area to the particle's geometric cross-sectional area.

The validity of Mie theory for our spray column system are based on several key assumptions : (1) scattering particles are perfectly spherical ; (2) particles are homogeneous with uniform optical properties and non-interacting, with negligible electromagnetic coupling between neighboring droplets ; and (3) the incident light is a plane wave with single scattering dominating. With typical droplet diameters of 50 micrometres and visible light wavelengths spanning 400–700 nanometres, the size parameter, m ranges from approximately 450 to 800. This places the spray system in the intermediate regime where Mie theory is fully applicable and interference effects between scattering and diffraction are fully accounted for.

The extinction phenomenon is defined as the combined effect of absorption and scattering. The complex Lorentz–Mie coefficients a_n and b_n are calculated to compute the extinction efficiency factor Q_{ext} and the scattering efficiency factor Q_{sca} .

$$a_n = \frac{\gamma^2 j_n(\gamma x) [x j_n(x)]' - j_n(x) [\gamma x j_n(\gamma x)]'}{\gamma^2 j_n(\gamma x) [x h_n^{(1)}(x)]' - h_n^{(1)}(x) [\gamma x j_n(\gamma x)]'} \quad (13)$$

$$b_n = \frac{j_n(\gamma x) [x j_n(x)]' - j_n(x) [\gamma x j_n(\gamma x)]'}{j_n(\gamma x) [x h_n^{(1)}(x)]' - h_n^{(1)}(x) [\gamma x j_n(\gamma x)]'} \quad (14)$$

where x is the size parameter, γ is the refractive index ratio, $h_n(z) = j_n(z) + i y_n(z)$, and $j_n(z)$ (resp. $y_n(z)$) refers to the Bessel function of the first kind of order n and argument z .

$$Q_{\text{ext}} = \frac{2}{x^2} \sum_{n=1}^{n_{\text{max}}} (2n+1) \text{Re}(a_n + b_n) \quad (15)$$

$$Q_{\text{sca}} = \frac{2}{x^2} \sum_{n=1}^{n_{\text{max}}} ((2n+1) (|a_n|^2 + |b_n|^2)) \quad (16)$$

$$n_{\text{max}} = x + 4x^{1/3} + 2 \quad (17)$$

where $\text{Re}(z)$ represents the real part of the complex quantity z . In practical computations, the infinite series is approximated at a finite order n_{max} .

The extinction coefficient β and the scattering coefficient σ is given by following classical Mie theory ([Bohren and Huffman, 1983](#)) :

$$\beta(z) = Q_{\text{ext}} n(z) A_g \frac{X_i + X_{d,i}}{X_{\text{ref}}}, \quad (18)$$

$$\sigma(z) = Q_{\text{sca}} n(z) A_g \frac{X_{d,i}}{X_{\text{ref}}}, \quad (19)$$

where Q_{ext} and Q_{sca} are the extinction and scattering efficiencies (dimensionless), $n(z)$ is the droplet or particle number density as a function of height z [m^{-3}], $A_g = \pi d^2/4$ is the geometric cross-sectional area of a particle of diameter d , X_i and $X_{d,i}$ are the active and dead biomass concentrations in stage i and X_{ref} is a reference biomass concentration used for non-dimensional scaling.

The equation for β and σ from [Bohren and Huffman \(1983\)](#) are adapted to evaluate the effect of biomass concentration on light attenuation. The biomass scaling factor $\frac{X_i}{X_{\text{ref}}}$ and $n(z)$ quantify the change in extinction β and scattering σ coefficients with increasing biomass concentration. These coefficients are applied to the radiative transfer equations in section [3.5.2](#), determining how light intensity attenuates as it propagates through layers of increasing droplet density.

3.5.2. Two-Flux Approximation

The PBR is modeled as a cylinder, therefore light incident is uniform along the circumference, which allows the full radiative transfer equation to be simplified to a 1D problem. The two-flux approximation defined by [Tseng and Viskanta \(2007\)](#) reduces the radiative field to two coupled ordinary differential equations describing forward flux F_X^+ and backward flux F_X^- with no reflection at the boundaries. The equations governing radiative transfer in the positive (+) and negative (-) directions are formulated as :

$$\frac{1}{\sqrt{3}} \frac{dF^+}{dr} = -\beta F^+ + \sigma(fF^+ + bF^-) + \kappa E(T) \quad (20)$$

$$-\frac{1}{\sqrt{3}} \frac{dF^-}{dr} = -\beta F^- + \sigma(fF^- + bF^+) + \kappa E(T) \quad (21)$$

where

- $\beta = \kappa + \sigma$ - extinction coefficient
- $f = 1 - b$ - forward scattering fraction
- b - back-scattering fraction
- κ - absorption coefficient
- σ - scattering coefficient

The water-spray curtain is approximated as a plane with a uniform layer of droplets, which matches the 1D and homogeneous assumptions underlying the two-flux model. This approximation is valid for a spray column when scattering is treated as linearly anisotropic, determined by fixed forward-scattering f and back-scattering fractions b . Since water droplets and microalgal cells exhibit strongly forward scattering, f is fixed at 0.15 while b is 0.85 throughout the simulation.

Boundary Conditions

$$F^+ = F^0 \quad \text{at } x = 0 \quad (22)$$

$$F^- = 0 \quad \text{at } x = L \quad (23)$$

Boundary conditions specify incident intensity at the column surface ($F_x^+ = F_x^0$) and zero flux at the bottom ($F_x^- = 0$). Radiative emission by droplets is negligible as temperature is below saturation temperature of water. In Block 1, the two-flux equations are solved independently for each layer using local extinction and scattering coefficients computed in section [3.5.1](#), yielding the downwelling irradiance available for photosynthesis at each radial position.

3.6. Growth Kinetics

The model by [Nikolaou et al. \(2015\)](#) predicts growth kinetics by considering dynamics of photosystem II states (RCII) and the non-photochemical quenching activity (NPQ). These fluxes can then be related to the growth rate by equations compiled by [Bernardi et al. \(2015, 2016\)](#).

$$\mu = \left(I\sigma\nu \frac{A\eta_P}{1 + \eta_D + \eta_{qE}\alpha + A\eta_P + C\eta_I} - R \right) \theta \quad (24)$$

$$\dot{A} = -I\sigma_{PS2}A + \frac{B}{\tau} \quad (25)$$

$$\dot{B} = I\sigma_{PS2}A - \frac{B}{\tau} + k_rC - k_d\sigma_{PS2}IB \quad (26)$$

$$\dot{C} = -k_rC + k_d\sigma_{PS2}IB \quad (27)$$

$$\dot{\alpha} = \xi \left(\frac{I^n}{I_{qE}^n + I^n} - \alpha \right) \quad (28)$$

$$\sigma_{PS2} = \frac{\sigma\eta_P}{N(1 + \eta_D + \eta_{qE}\alpha + \eta_P)} \quad (29)$$

TABLE 2: Growth kinetic parameters [\(Nikolaou et al., 2015\)](#)

Parameter	Value	Units
Turnover time, τ	5×10^{-3}	s
Damage rate, k_d	6.41×10^{-7}	—
Repair rate, k_r	5.55×10^{-5}	s^{-1}
NPQ rate, ξ	5.95×10^{-2}	s^{-1}
Hill coefficient, n	2.26	—
Photoproduction, η_P	1.14×10^{-1}	—
Thermal decay, η_D	5.00×10^{-2}	—
qE-quenching, $\bar{\eta}_{qE}$	1.98×10^{-1}	—
qI-quenching, η_I	7.87×10^{-1}	—
Respiration, R	1.78×10^{-5}	$\text{g C g}^{-1} \text{ chl s}^{-1}$
Stoichiometry, ν	4.0	e^-/C

The growth rate μ [s^{-1}] in equation [24](#) is proportional to the chlorophyll-to-carbon quota θ [$\text{g}_{\text{chl}} \text{g}_{\text{C}}^{-1}$]. It is also determined by the incident light intensity I [$\mu\text{E m}^{-2} \text{s}^{-1}$], the effective absorption cross-section σ [$\text{m}^2 \text{g}_{\text{chl}}^{-1}$], and the respiration coefficient R [$\text{g}_{\text{C}} \text{g}_{\text{chl}}^{-1} \text{s}^{-1}$], with ν acting as the stoichiometric factor that associates electron transport with carbon fixation. The transition of PSII reaction centres (RCII) between active state A , the occupied state B , and the damaged state C as function of irradiance I are

denoted in equation (25) - (27) following Han (2002). The transition $A \rightleftharpoons B$ computes the photoproduction, which depends on the PSII absorption cross-section σ_{PS2} [$\text{m}^2 \mu\text{E}^{-1}$] and the RCII turnover time τ [s]. Photoinhibition corresponds to the transition $B \rightleftharpoons C$, which is governed by the dimensionless damage parameter k_d and the repair rate constant k_r [s^{-1}]. The change of acclimation parameter α is modelled as a first-order relaxation toward a light-dependent steady state and expressed as shown in equation (28). Equation (29) defines the relationship between σ_{PS2} to the number of RCII units N [$\mu\text{E g}_{\text{chl}}^{-1}$], σ and α .

Microalgae undergo photoacclimation in response to chronic changes in light environment, pigment antenna sizes and photosynthetic reaction center abundance (Bernard, 2011). The acclimation state is characterized by the acclimation irradiance I_g , a time-integrated measure of growth light history that evolves according to :

$$\dot{I}_g = \delta(I - I_g)\mu \quad (30)$$

where δ is a dimensionless acclimation rate coefficient, I is the instantaneous irradiance, and μ is the growth rate. This ensures the acclimation state tracks prevailing light conditions, with high irradiance triggering photoprotective responses and low irradiance promoting light-harvesting enhancement.

The physiological adjustments associated with photoacclimation are captured through power-law relationships relating key variables to I_g . The variation in pigment content and the abundance of Photosystem II reaction centers are described by (MacIntyre et al., 2002; Fisher et al., 1996; Nikolaou et al., 2016b) :

$$\theta = \theta_0(I_g)^\kappa \quad (31)$$

$$N = N_0(I_g)^\lambda \quad (32)$$

where θ is the chlorophyll-to-carbon quota, N is the RCII abundance, and θ_0 , N_0 , κ , λ are acclimation parameters. The carotenoid-to-chlorophyll ratio also adjusts due to photoacclimation impacts on excess energy dissipation via non-photochemical quenching, captured through (Simionato et al., 2011) :

$$I_{qE} = I_{qE,0} + \epsilon I_g \quad (33)$$

with corresponding acclimation parameters I_0 and ϵ .

When exposed to sustained low light, cells increase chlorophyll content to maximize photon capture. Conversely, under persistent high light, cells downregulate pigment content and increase reaction center abundance to enhance energy dissipation and prevent photoinhibition.

The rationale behind the choice of the acclimation rules in equations (31) - (33) is that they provide sufficient

TABLE 3: Estimated acclimation and respiration parameter values (Nikolaou et al., 2015)

Parameter	Value	Units
Acclimation rate, δ	1.38×10^{-1}	—
Reference chl quota, θ_0	3.60	$\text{g chl g}^{-1} \text{ C}$
Chl exponent, κ	-1.89×10^{-1}	—
Reference RCII, N_0	1.71	$\mu\text{mol g}^{-1} \text{ chl}$
RCII exponent, λ	1.22×10^{-1}	—
NPQ intercept, $I_{qE,0}$	8.13×10^2	$\mu\text{mol m}^{-2} \text{ s}^{-1}$
NPQ slope, ϵ	4.48×10^{-1}	—

complexity for accurate prediction of a set of experimental data for the associated photoacclimation parameters θ_0 , N_0 , κ , and λ with statistically-confident estimates (Nikolaou et al., 2015). The overall growth model (24) - (33) accounts for photoproduction, photoregulation, photoinhibition and photoacclimation all together, and its dynamics span many timescales ranging from milliseconds to days.

3.7. System Dynamics Coupling

Block 1 represented by section 3.5 solves the two-flux radiative transfer equations for each layer n to determine the spatially-resolved irradiance. The circumferential LED illumination yields a radially-symmetric light field. The photosynthetic growth rate varies radially across the layer cross-section ; thus, a radially-integrated growth rate $\bar{\mu}_i$ is computed for each layer :

$$\bar{\mu}_i = \frac{2}{R^2} \int_0^R \mu(r) r dr \quad (34)$$

where r is the radial coordinate. This averaged growth rate characterizes the mean photosynthetic productivity of the biomass population in layer n and serves as input to the biomass balance.

A CSTR mass balance to each layer, incorporating downward transport from the upstream layer and biological transformations. For active biomass concentration X_n :

$$\frac{dX_n}{dt} = D(X_{n-1} - X_n) + \bar{\mu}_n X_n \quad (35)$$

$$\frac{dX_{d,n}}{dt} = D(X_{d,n-1} - X_{d,n}) + R_{\text{resp}} X_n \quad (36)$$

with inlet boundary condition $X_0 = X_{\text{inlet}}$ and outlet condition at layer N with possible harvesting extraction.

Column-wide averages are computed from the layer-resolved states :

$$\bar{X}_{\text{total}}(t) = \frac{1}{N} \sum_{n=1}^N X_n(t) \quad (37)$$

$$\bar{\mu}_{\text{total}}(t) = \frac{1}{N} \sum_{n=1}^N \mu_n(t) \quad (38)$$

4. Results and Discussion

4.1. Light Attenuation

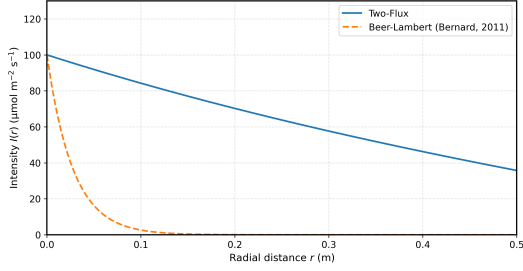


FIGURE 6: Beer Lambert vs Two-Flux Approximation

Light attenuation was analysed using both Beer-lambert law and two-flux approximation under same operational parameters. The Beer-lambert model describes the light transport if the system was a bulk liquid, while two-flux approximation accounts for the light attenuation within the spray column configuration. The graph shown in Figure 6 denotes that light decays faster with Beer-lambert model than two-flux approximation model. Beer-lambert model assumes that there is purely absorption in the medium, so photons are lost permanently from every interaction. Two-flux approximation represents scattering as a redistribution of photons between forward and backward fluxes. A portion of lost photons are redirected back to the forward flux, increasing effective optical path length travelled through the reactor, and consequently slowing down the process of light decay.

4.2. Biomass Change

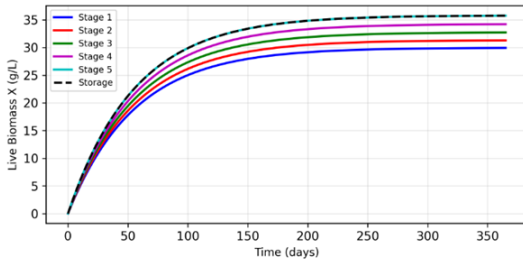


FIGURE 7: Living Biomass Change Over Time

The live biomass concentration exhibits an asymptotic growth profile over the one-year simulation period. From the graph, the live biomass in each layer increases steeply due to rapid accumulation at the start until it reaches plateau around 250 days. During the early simulation period, the net specific growth rate is exceeding the decay and dilution rate, therefore biomass accumulates. The plateau is attributed to the continuous liquid medium in the CSTR, which the system achieves steady rate as the growth and removal rate cancel out. Microalgae cells experience self-shading over time due to increasing biomass concentration

as more cells that scatter light are present along the optical path (Saccardo et al., 2022). Light attenuation amplified so less light is available to penetrate into the deeper region of the PBR. The biomass concentration increases across successive layers, which the last layer in the PBR and storage tank achieve similar biomass concentration. This indicates that there is no growth rate in the tank as no new biomass is produced. Predicted biomass concentrations exceed those of open-raceway systems 3.0 g/L as reported by Nikolaou et al. (2016b) by over an order of magnitude, highlighting the productivity advantage of enclosed PBR designs.

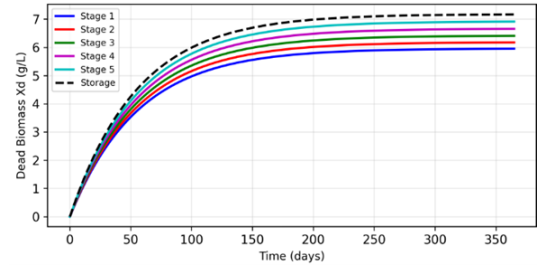


FIGURE 8: Dead Biomass Change Over Time

The dead biomass concentration exhibits the same asymptotic growth profile as the live biomass concentration over the one-year simulation period. Rapid accumulation as denoted by the steep curve in early simulation days is a result of dead biomass from the PBR transported to the tank while continuously producing new dead biomass due to respiration of cells. In the dark storage tank, photosynthesis is inhibited and the cells rely on respiration of internal reserves for maintenance process, thereby resulting in decay of biomass (Adamson, 2015). Over long periods, alternating light-dark cycles can convert a small fraction of energy-limited cells into dead cells.

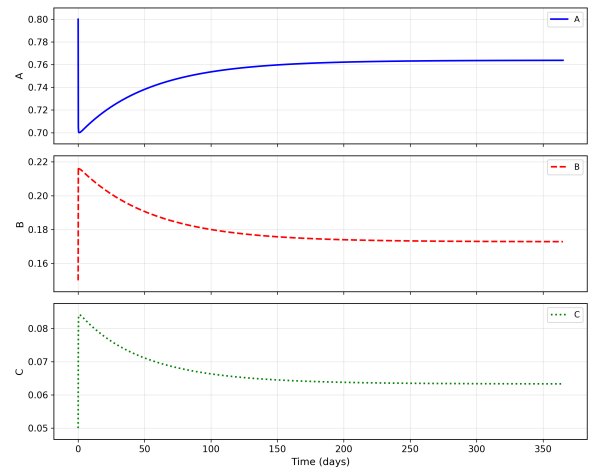


FIGURE 9: Photosystems A, B and C Over Time

Active state (state A) in the PBRs experience a sharp drop from 0.80 to 0.70 in the early simulation period fol-

lowing a slow increase to its steady state, while the value of state A in the storage tank decreases from 1 to 0.92 approximately before achieving steady state. A sharp drop of state A from the initial value is due to PS2 reaction centres in the cells receiving photons from the irradiance field and become closed. Light decays radially over longer timescales due to increasing biomass cells in the PBRs resulted in lower amount of active reaction centres. Therefore, a gradual increase in state A reveals that photoacclimation and PS2 repair process occurs reducing the fraction of closed reaction centres (Wu et al., 2012). The absence of light in the storage tank suggests that PS2 is exposed to dark decay. The initial peak in state B is driven by the excitation of open centres in state A. Subsequently, the gradual decrease arises from photoacclimation as the lifetime for occupied PS2 centres is shorter. The sudden escalation of state C reflects the high incident irradiance, causing a larger fraction of photosystem to experience photodamage. The following steady drop is due to lower photodamage as a result of lower light irradiance.

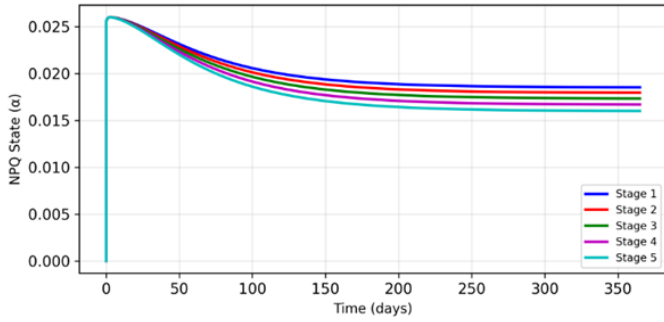


FIGURE 10: NPQ Over Time

The NPQ α exhibits a steep rise, reflecting rapid activation of non-photochemical quenching in response to exposure of light irradiance. After the peak, the α decrease progressively over 300 days until it slowly reaches a non-zero steady state value. It is evident that a marginal portion of absorbed light is dissipated as heat to protect the cells from photodamage.

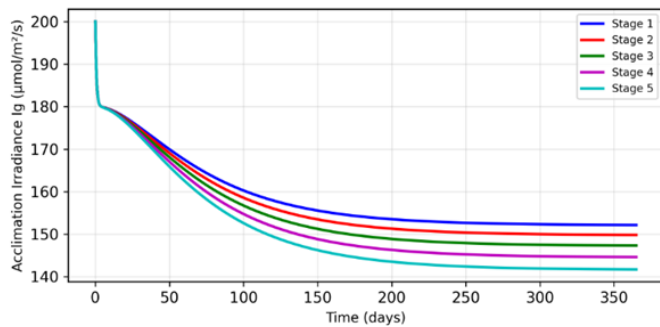


FIGURE 11: NPQ Over Time

Acclimation irradiance shows a monotonic abrupt decrease from its initial values before approaching a quasi-

steady state. This is triggered by the cell adjustment to a reduced light level within the reactor cascade. As culture experiences self-shading and lower irradiance in deep regions, microalgae cells shift towards a lower acclimation of irradiance, which promotes higher pigment content and effective absorption cross section per cell. This pattern enhances light-harvesting efficiency under light-capture growth conditions, thereby improving photosynthetic activity.

4.3. Effect of Dilution Rate and Number of Stages on Biomass Production

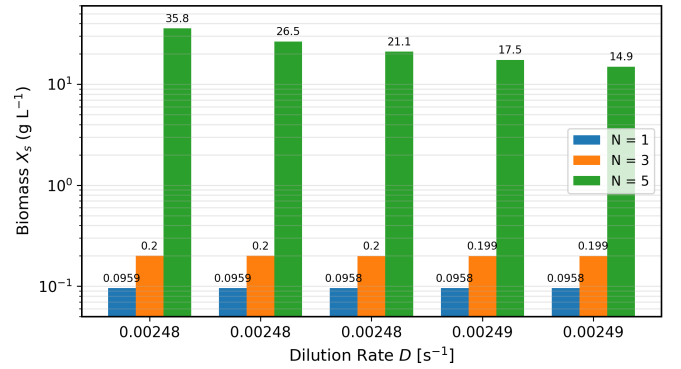


FIGURE 12: Effect of Dilution Rate and Number of stages on Final Biomass Production

As described in figure 12, the final biomass concentration increases with smaller dilution rates. The physical description of dilution rate is how fast material is being circulated through the system. When dilution rates are small, residence time in a layer is increased. Thus, the algae spends increased time in light zones, so there is more growth occurring.

The increase in biomass with finer spatial discretization ($N = 1$ to $N = 5$) reflects the importance of resolving photoacclimation dynamics across the light gradient. Coarse discretization treats the reactor as spatially uniform, underestimating productivity by averaging physiological responses across regions with different light availability. Refined discretization allows independent acclimation per layer. This spatial heterogeneity drives the observed 3.8-fold increase in biomass, highlighting the necessity of adequate layer resolution for accurate reactor productivity predictions.

5. Conclusion

The proposed model describes the first comprehensive, integrated model for a sb-PBR. The model couples light distribution events, spray droplet dynamics and photosynthetic growth kinetics into compiled framework. This multi-physics characterization of spray-based systems fills a methodological gap in PBR modelling.

The theoretical predictions produced by this framework suggests increased biomass production compared to

conventional PBRs. These results can be substantiated by the light irradiance dynamics in this model where it is assessed to have a better light utilisation performance in comparison to classical Beer-Lambert theory based models. Future endeavours in this work should focus on pilot-scale experiments to validate model predictions and calibration of physiological parameters of operation. This can be further worked into systematic optimisation for identifying optimal configurations for commercial-scale applications.

6. Acknowledgements

The authors would like to thank David Danaci and Alberto Saccardo for their support and guidance during the duration of the project.

References

- Adamson, L., 2015. The effect of storage on algae biomass composition based on the model organism *Neochloris oleoabundans*.
- Agbebi, T., 2019. Design and Characterization of a Miniature Photobioreactor for Microalgae Culture. Ph.D. thesis. University College London.
- Benner, P., Meier, L., Pfeffer, A., Krüger, K., Oropeza Vargas, J., Weuster-Botz, D., 2022. Lab-scale photobioreactor systems : principles, applications, and scalability. *Bioprocess and Biosystems Engineering* 45, 791–813. doi: [10.1186/s12934-022-01234-9](https://doi.org/10.1186/s12934-022-01234-9)
- Bernard, O., 2011. Hurdles and challenges for modelling and control of microalgae for CO mitigation and biofuel production. *Journal of Process Control* 21, 1378–1389. doi: [10.1016/j.jprocont.2011.07.012](https://doi.org/10.1016/j.jprocont.2011.07.012)
- Bernard, O., Mairet, F., Chachuat, B., 2016. Modelling of microalgae culture systems with applications to control and optimization. *Advances in Biochemical Engineering/Biotechnology* 153, 59–87. doi: [10.1007/10_2014_287](https://doi.org/10.1007/10_2014_287)
- Bernardi, A., Nikolaou, A., Meneghesso, A., Morosinotto, T., Chachuat, B., Bezzo, F., 2015. A framework for dynamic modeling of PI curves in microalgae, in : *Computer Aided Chemical Engineering*, pp. 2483–2488.
- Bernardi, A., Nikolaou, A., Meneghesso, A., Morosinotto, T., Chachuat, B., Bezzo, F., 2016. High-fidelity modelling methodology of light-limited photosynthetic production in microalgae. *PLoS ONE* 11, e0152387. doi: [10.1371/journal.pone.0152387](https://doi.org/10.1371/journal.pone.0152387)
- Bird, R., Stewart, W., Lightfoot, E., 2007. Transport phenomena .
- Bohren, C.F., Huffman, D.R., 1983. Absorption and Scattering of Light by Small Particles. Wiley, New York.
- Buchmann, N., Atkinson, C., Soria, J., Sheridan, J., 2014. Tomographic particle image velocimetry investigation of the flowfield in shallow reservoir swimming strokes. *Journal of Biomechanics* 47, 1632–1639. doi: [10.1007/s11095-014-1391-6](https://doi.org/10.1007/s11095-014-1391-6)
- Chanquia, S., Vernet, G., Kara, S., 2021. Photobioreactors for cultivation and synthesis : specifications, challenges, and perspectives. *Engineering in Life Sciences* 22, 712–724. doi: [10.1002/elsc.202100070](https://doi.org/10.1002/elsc.202100070)
- Fathurrahman, L., Siti Hajar, A., Wan Nur Sakinah, D., Nurhazwani, Z., Ahmad, J., 2013. Flashing light as growth stimulant in cultivation of green microalgae, *Chlorella* sp. utilizing airlift photobioreactor. *Pakistan Journal of Biological Sciences* 16, 1517–1523.
- Fisher, T., Minnaard, J., Dubinsky, Z., 1996. Photoacclimation in the marine alga *Nannochloropsis* sp. (eustigmatophyte) : A kinetic study. *Journal of Plankton Research* 18, 1797–1818. doi: [10.1093/plankt/18.10.1797](https://doi.org/10.1093/plankt/18.10.1797)
- Geng, X., Qu, J., Yin, X., Wang, Q., Qin, S., 2025. Progress in therapeutic applications of microalgae and their bioactive compounds. *Algal Research* 91, 104226. doi: [10.1016/j.algal.2025.104226](https://doi.org/10.1016/j.algal.2025.104226)
- Khan, M., Shin, J., Kim, J., 2018. The promising future of microalgae : current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. *Microbial Cell Factories* 17. doi: [10.1186/s12934-018-0985-x](https://doi.org/10.1186/s12934-018-0985-x)
- Klinthong, W., Yang, Y.H., Huang, C.H., Tan, C.S., 2015. A review : Microalgae and their applications in CO capture and renewable energy. *Aerosol and Air Quality Research* 15, 712–742. doi: [10.4209/aaqr.2014.11.0299](https://doi.org/10.4209/aaqr.2014.11.0299)
- MacIntyre, H., Kana, T., Anning, T., Geider, R., 2002. Photoacclimation of photosynthesis irradiance response curves and photosynthetic pigments in microalgae and cyanobacteria. *Journal of Applied Phycology* 38, 17–38. doi: [10.1046/j.1529-8817.2002.00003.x](https://doi.org/10.1046/j.1529-8817.2002.00003.x)
- Nikolaou, A., Bernardi, A., Meneghesso, A., Bezzo, F., Morosinotto, T., Chachuat, B., 2015. A model of chlorophyll fluorescence in microalgae integrating photoproduction, photoinhibition and photoregulation. *Journal of Biotechnology* 194, 91–99. doi: [10.1016/j.jbiotec.2014.11.025](https://doi.org/10.1016/j.jbiotec.2014.11.025)
- Nikolaou, A., Bernardi, A., Meneghesso, A., Chachuat, B., Morosinotto, T., Bezzo, F., 2016a. Multi-scale model of photosynthetic growth of microalgae in photobioreactors, in : *Computer Aided Chemical Engineering*, pp. 1–6. doi: [10.1016/B978-0-444-63428-3.50001-0](https://doi.org/10.1016/B978-0-444-63428-3.50001-0)
- Nikolaou, A., Hartmann, P., Sciandra, A., Chachuat, B., Bernard, O., 2016b. Dynamic coupling of photoacclimation and photoinhibition in a model of microalgae growth. *Journal of Theoretical Biology* 390, 61–72. doi: [10.1016/j.jtbi.2015.11.008](https://doi.org/10.1016/j.jtbi.2015.11.008)
- Razzak, S., Bahar, K., Islam, K., Haniffa, A., Faruque, M., Hossain, S., Hossain, M., 2023. Microalgae cultivation in photobioreactors : sustainable solutions for a greener future. *Green Chemical Engineering* 5, 418–439. doi: [10.1016/j.gce.2023.10.004](https://doi.org/10.1016/j.gce.2023.10.004)
- Saccardo, A., Bezzo, F., Sforza, E., 2022. Microalgae growth in ultra-thin steady-state continuous photobioreactors : assessing self-shading effects. *Frontiers in Bioengineering and Biotechnology* 10, 977429. doi: [10.3389/fbioe.2022.977429](https://doi.org/10.3389/fbioe.2022.977429)
- Schulze, P., Barreira, L., Pereira, H., Perales, J., Varela, J., 2014. Flashing light through oxygen-enriched air enhances growth in *Chlorella vulgaris*. *Biotechnology and Bioengineering* 111, 2477–2485. doi: [10.1002/bit.28992](https://doi.org/10.1002/bit.28992)
- Sforza, E., Simionato, D., Principale, G., Tredici, M., Bertucco, A., 2020. Computational fluid dynamics simulation of an industrial photobioreactor. *Applied Sciences* 9, 2843. doi: [10.3390/app9142843](https://doi.org/10.3390/app9142843)
- Simionato, D., Sforza, E., Corteggiani Carpinelli, E., Bertucco, A., Giacometti, G., Morosinotto, T., 2011. Acclimation of *Nannochloropsis gaditana* to different illumination regimes : Effects on lipids accumulation. *Bioresource Technology* 102, 6026–6032. doi: [10.1016/j.biortech.2011.02.055](https://doi.org/10.1016/j.biortech.2011.02.055)
- Tseng, C., Viskanta, R., 2007. Absorptance and transmittance of water spray/mist curtains. *International Journal of Heat and Mass Transfer* 50, 4852–4859. doi: [10.1016/j.ijheatmasstransfer.2007.02.013](https://doi.org/10.1016/j.ijheatmasstransfer.2007.02.013)
- Uyar, B., Ali, M., Ozer Uyar, G., 2024. Design parameters comparison of bubble column, airlift and stirred tank photobioreactors for microalgae production. *Bioprocess and Biosystems Engineering* 47, 195–209. doi: [10.1186/s12934-024-02437-5](https://doi.org/10.1186/s12934-024-02437-5)
- Wu, H., Roy, S., Alami, M., Green, B.R., Campbell, D.A., 2012. Photosystem II photoinactivation, repair, and protection in marine centric diatoms. *Plant Physiology* 160, 464–476. doi: [10.1104/pp.112.203067](https://doi.org/10.1104/pp.112.203067)

Fabrication and Characterisation of Matrimid 5218-CPOC-301 Mixed-Matrix Membranes

Alexander Turner, Sisi Guo

Department of Chemical Engineering, Imperial College London

December 11, 2025

1 Abstract

High-performance polymer membranes are playing an increasingly important role in improving molecular separation across chemical and environmental applications. Membrane-based water filtration and crude oil fractionation are gaining recognition as effective alternatives to traditional processes. In this study, composite membranes were fabricated using Matrimid 5218 polyimide mixed with the porous organic cage CPOC-301, with an optional polyacrylonitrile (PAN) substrate. The aim was to prepare and characterise these mixed-matrix membranes to assess the influence of cage incorporation on permeance and selectivity. Two experimental approaches were adopted: (i) pressure-driven permeation of methanol to determine solvent permeance, and (ii) transport of mixed dye solutions to evaluate separation. The morphology, surface structure, and textural properties of the membrane were characterised using scanning electron microscopy (SEM) and CO₂ adsorption analysis, while ultraviolet–visible spectroscopy (UV-Vis) and high-performance liquid chromatography (HPLC) were used to quantify dye transport and permeate composition. The membranes exhibited methanol permeance up to 32.52 mL m⁻² min⁻¹ bar⁻¹ and 100% Vitamin B12 rejection after stabilisation. Overall, this study outlines a clear methodology for fabricating and evaluating Matrimid-5218-CPOC-301 composite membranes, and demonstrates that cage incorporation leads to measurable changes in morphology and liquid-phase transport.

2 Introduction

Membrane-based separation technologies have become increasingly significant across chemical and environmental engineering due to their potential to provide energy-efficient alternatives to traditional separation processes [1]. Applications such as water purification [2], wastewater treatment, and even crude oil fractionation [3] demonstrate the capability of membranes to reduce energy demand, minimise solvent use, and enable continuous processing. As global interest in sustainable and low-energy separations grows, the development of high-performance membrane materials capable of achieving both high permeance and strong molecular selectivity has become a major research priority [4].

Polymer membranes offer mechanical robustness, scalability, and can be easily processed, making them well suited for liquid-phase separations. However, their performance is often limited by a permeability–selectivity trade-off. Matrimid 5218, a commercially available polyimide, is widely used as a benchmark material due

to its thermal stability [5], solvent resistance, and established performance in gas separation and organic solvent nanofiltration. Nevertheless, its relatively low microporosity and dense packing can limit solvent transport, motivating the development of composite or modified membrane architectures to enhance free volume and transport efficiency.

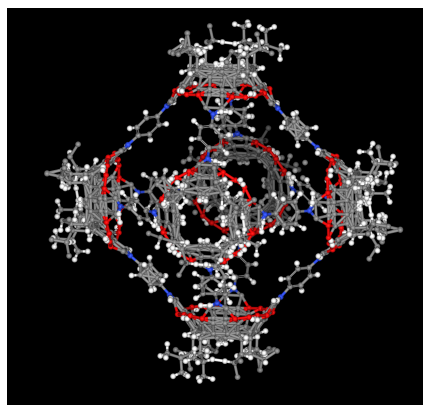


Figure 1: Molecular structure of CPOC-301

Mixed-matrix membranes (MMMs) have emerged as an effective strategy to overcome these limitations by incorporating porous fillers into polymer matrices to improve free volume, modify polymer packing, and introduce additional transport pathways[6]. Porous organic cages (POCs) are a promising class of fillers due to their well-defined, rigid molecular cavities and inherent solubility in common organic solvents. Soluble porous organic cages have previously been shown to mix uniformly with polymers and enable high loading MMMs [7]. CPOC-301, the cage material employed in this study, dissolves readily in solvents such as chloroform and dichloromethane, enabling homogeneous mixing with Matrimid and allowing MMMs with very high cage loadings—up to 80 wt% and even preparation of pure-cage membranes. The cage has an intrinsic cavity and open windows of approximately 0.7 nm, providing accessible microporosity and additional transport channels for small molecules. As a result, increasing cage loading can significantly increase permeance while keeping its size-based selectivity, allowing tailored rejection of larger dye molecules.

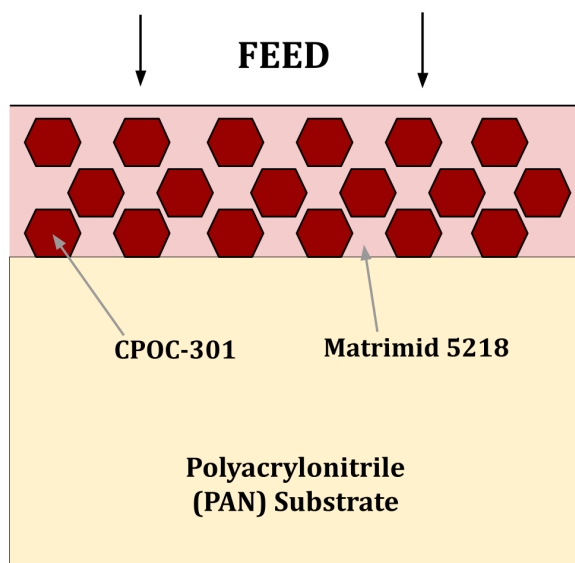


Figure 2: Uniform Cage Distribution of Matrimid 5218 / CPOC-301 on PAN Substrate

In addition to filler properties, membrane structure influences separation performance. Polyacrylonitrile (PAN) is commonly used as a porous support material due to its mechanical strength and resistance to organic solvents, providing a stable base for thin selective layers[8]. The combination of a robust PAN support with a highly loaded Matrimid/POC selective layer forms a structure that can be adapted not only for organic solvent nanofiltration but also for gas separation, desalination, and a range of advanced

membrane applications.

This study aims to fabricate mixed-matrix membranes composed of Matrimid 5218 and CPOC-301, with and without a PAN support, and to characterise their structural and transport properties. Methanol permeation experiments under applied pressure were carried out to evaluate solvent transport behaviour, while Vitamin B12 rejection tests were conducted to examine molecular selectivity and membrane stability. Characterisation techniques including scanning electron microscopy (SEM), CO₂ adsorption isotherms, ultraviolet–visible spectroscopy (UV-Vis), and high-performance liquid chromatography (HPLC) were employed to analyse morphology, porosity, and dye transport. Together, these methods provide insight into how porous organic cages influence the structure and performance of Matrimid-based mixed-matrix membranes for liquid-phase separations.

3 Methodology

3.1 Membrane Preparation

3.1.1 Preparation of Matrimid/CPOC-301 Coating Solution

A Matrimid/CPOC-301 casting solution in chloroform was prepared to form the selective layer of the mixed-matrix membrane (MMM). Matrimid 5218 and CPOC-301 were weighed so that CPOC-301 represented 50 wt% of the combined mass of cage and polymer in a 0.5wt% chloroform solution. The solids were dissolved in chloroform at room temperature followed by mixing until a homogeneous solution without visible particulates was obtained. This low-viscosity Matrimid/CPOC-301 solution was then used as the coating solution to cast unsupported membranes and to form the mixed-matrix selective layer on PAN supports.

3.1.2 Screen Coated Matrimid–CPOC-301 Mixed-Matrix Membranes

PAN support membrane was cut into squares with sides of approximately 4 cm. A Matrimid/CPOC-301 solution was applied to a porous PAN substrate via screen coating to produce a PAN-supported composite membrane. The membrane edges were secured to a flat glass plate with tape to prevent wrinkling. 3ml of Matrimid/CPOC-301 coating solution was applied to the PAN surface and swiftly and uniformly distributed using a glass rod to create a thin liquid layer over the surface. The coated film was subsequently cov-

ered at ambient temperature to ensure gradual evaporation of chloroform, followed by oven drying at 80 °C to remove the residual solvent.

3.1.3 Spin Coated Matrimid–CPOC-301 Mixed-Matrix Membranes

The PAN supported composite membrane was prepared using spin coating, where 0.5wt% Matrimid/CPOC-301 solution was coated onto a porous PAN substrate.

Firstly, a PAN support sheet larger than the final membrane testing area was cut for spin coating. The wetted PAN support sheet was then fixed to a metal chuck using ordinary transparent tape and clamped onto the spin coater chuck using vacuum suction. A 3mL of Matrimid/CPOC-301 solution was dropped onto a small central area of the fixed support material, ensuring adequate coverage. To avoid streaks in the final film, attempts were made to prevent deposition at a single central point.



Figure 3: Spin coater used to apply the Matrimid/CPOC-301 selective layer onto PAN support membranes

The spin coating process consisted of two steps: an initial spin of 500 rpm for 18 seconds, followed by 2000 rpm for 60 seconds. This gradual increase in speed ensured that the liquid film reached a uniform state before most of the solution is spun off, promoting uniform film thickness and preventing immediate solution run-off. After completion, the metal plate was removed from the chamber, and the PAN support material was carefully peeled off the tape, leaving a thin layer.

3.1.4 Drying and Solvent Activation / Conditioning

The coated support membrane was then kept at ambient temperature under a lid for a minimum of 24 hours to allow chloroform to evapo-

rate gradually. In this way, a continuous selective Matrimid/CPOC-301 layer was formed on the porous PAN support. After this initial ambient drying step, all Matrimid/CPOC-301 membranes were further dried in an oven for 24 h at 80 °C to remove residual chloroform and stabilise the polymer–cage network.

Prior to any transport experiments, the membrane was activated in a pressure vessel using hexane. This removed residual solvent, opened and stabilised the pore structure, resulting in a membrane with consistent performance and good wetting, facilitating fair performance comparisons. The selective layer, Matrimid/CPOC-301, faced the feed side. The vessel pressure was increased to 10 bar, and hexane was allowed to flow through the membrane for at least one hour while permeate was continuously drained. Pressurized hexane flushing was used to remove residual chloroform and other solvents, saturating the pores with low-surface-tension hexane and physically activating the selective layer and the polyacrylonitrile (PAN) support layer.

Before testing methanol permeability or dye rejection, the system was allowed more time to eliminate residual hexane. After reducing vessel pressure, a fresh solvent was injected through the membrane and any permeate was removed until no hexane remained. This prevented the solvent from affecting transport measurement accuracy and uniformity.

3.2 Methanol Permeation Experiments

3.2.1 Pressure-Step Methanol Tests

Tests investigated the variation of permeance with pressure and the stability and compactness of the PAN-supported Matrimid/CPOC-301 membrane within a 2–20 bar operational range. Circular strips with a diameter of 1.0 cm (effective membrane area = 0.79 cm²) were cut from the sheets using a hollow die, specifically from the most uniform regions. A metallic support plate was located at the base of a 500 mL stainless-steel dead-end pressure vessel, where each strip was placed. A 10 mm stainless steel flow reducer was positioned above the membrane to expose the active area and provide a seal.

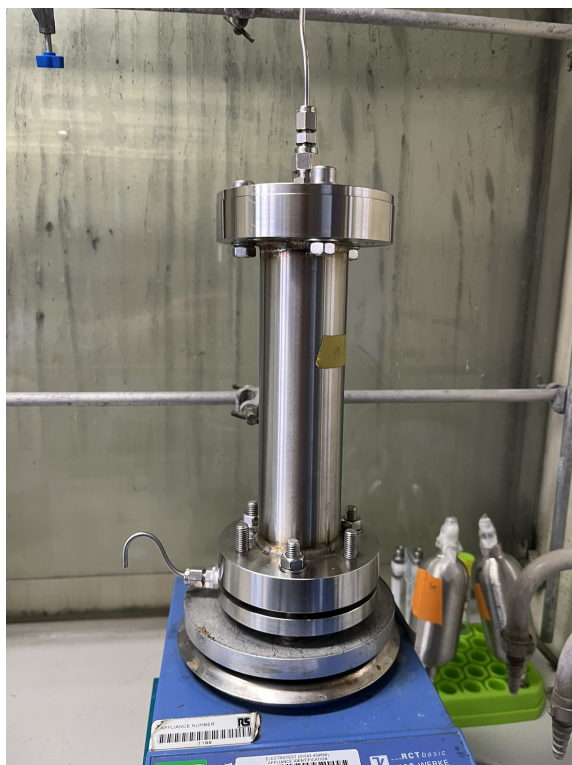


Figure 4: Stainless-steel high-pressure filtration vessel used for methanol permeation and dye rejection experiments

In all experiments, the magnetic stir bar was positioned approximately 10 mm above the flow reducer. Stirring was crucial in the dye rejection tests to prevent the accumulation of rejected solutes on the membrane surface.

Methanol (0.791 g mL^{-1}) was poured into the pressure vessel, and the transmembrane pressure was appropriately adjusted via the air cylinder's pressure regulator. Flux measurements were performed at pressures of 2, 5, 10, 13, 18, and 20 bar. At each pressure, the system reached equilibrium until steady-state flow was observed. Permeate samples were taken at predetermined time intervals: every 60 minutes at 2 and 5 bar, and every 30 minutes at 10, 13, 18, and 20 bar.

The net mass of the permeate was recorded using an analytical balance, and its volume was converted to mass using the density of methanol. The permeance was then calculated using the following formula:

$$\text{Permeance} = \frac{(m \text{ (g)} \rho \text{ (g mL}^{-1})) \times 10^{-3}}{A \text{ (m}^2) (t \text{ (min)} 60) \Delta P \text{ (bar)}}$$

To improve the accuracy of the measurements, three permeate samples were collected at each pressure step. The permeance was calculated separately for each sample using the above equation, and the arithmetic mean of the three values was taken as the reported permeance at that pressure.

3.3 Structural and Textural Characterisation

3.3.1 Scanning Electron Microscopy (SEM)

The surface and cross-sectional morphology of the film were observed using scanning electron microscopy (SEM). First, the film strip was thoroughly dried at room temperature to remove residual solvent. After manual cutting, the cross-sectional structure of the cut sample was fixed to an aluminium sample stage using conductive carbon tape, ensuring the cross-sectional direction of the membrane composed of Matrimid/CPOC-301 and the PAN support layer was perpendicularly exposed to the electron beam. Then, a thin gold layer was deposited via gold sputtering (Emitech K575X) to improve conductivity.

SEM imaging was performed on a Zeiss AU-RIGA CrossBeam field-emission SEM equipped with an InLens secondary electron detector. The accelerating voltage was set to 5.0 kV and magnifications up to approximately $20\,000\times$ were used. The InLens mode was primarily employed to capture images for high-resolution, surface-sensitive contrast of the Matrimid/CPOC-301 selective layer. The membrane surface and cross-section were concurrently analysed to determine the continuity and thickness of the selected layer, as well as to qualitatively assess the dispersion of CPOC-301 within the Matrimid matrix.

3.3.2 CO₂ Adsorption Isotherms

CO₂ adsorption-desorption isotherms were measured using a Micromeritics 3Flex surface area and porosity analyser to quantify the membrane's microporosity and CO₂ uptake. Samples were prepared by cutting dense Matrimid 5218 membranes, Matrimid/CPOC-301 mixed matrix membranes with different cage loadings (C10, C30, C50, and C80), pure CPOC-301 membranes, and CPOC-301 crystalline powder into small pieces, placing them in glass sample tubes, and recording their mass.

Before being analysed, all samples were first pretreated by degassing overnight under dynamic vacuum at 150°C to remove any absorbed moisture and residual solvent. The samples were then cooled to room temperature while still under vacuum, and the sample tubes were sealed to keep water and CO₂ from re-adsorbing. The activated samples were then put into the 3Flex's analysis port and put in an ice-water bath for a while to let them fully adjust to the bath temperature. The temperature stayed at 273 K during the measurement process.

The physisorption isotherm mode was selected, with CO₂ as the gas. A relative pressure range was set in the low-pressure region ($P/P_0 = 0.000\text{--}0.03$). After starting the analysis, 3Flex gradually increased the pressure of CO₂ in the sample tube, automatically waiting for equilibrium at each pressure point before recording the adsorption amount, until the set maximum P/P_0 was reached. Once the maximum relative pressure was reached, the desorption branch was recorded by gradually decreasing the pressure, and any hysteresis was identified.

After the test, the adsorption branch in the low-pressure region was selected for data analysis: the CO₂ adsorption amount at low P/P_0 was used to estimate the micropore volume and compare the porosity of Matrimid 5218, CPOC-301, and Matrimid/CPOC-301 hybrid matrix composites.

3.4 Dye Rejection Tests

3.4.1 Feed Solution Preparation

Vitamin B12 (cyanocobalamin) was selected as the model dye to evaluate the retention performance of the PAN-supported Matrimid/CPOC-301 membrane. The stock solution was made by dissolving vitamin B12 powder in methanol, resulting in a concentration of 0.01wt% vitamin B12 relative to the solution. The combination was agitated at room temperature until the solid was entirely dissolved, and the resultant stock solution was kept in the dark to prevent photodegradation.

3.4.2 Filtration Protocol

To evaluate the long-term dye rejection performance of the membrane, filtration experiments were conducted for seven consecutive days. Each day, the membrane was fed with a prepared Vitamin B12/methanol solution at a transmembrane pressure of 19 bar for approximately 2 hours. During each 2-hour run, permeate was continuously collected. At the end of each day's run, one permeate sample and one corresponding feed sample were analysed. The daily flux and rejection rate were then calculated based on these samples to track the membrane performance over time.

3.4.3 Quantification by UV-Vis and HPLC

The concentration of Vitamin B12 in the methanol permeate solution was quantitatively analysed using ultraviolet-visible (UV-Vis) spectrophotometry. A NanoDrop One Microvolume UV-Vis spectrophotometer was used, with

methanol as a blank solution, to collect UV-Vis absorption spectra in the wavelength range of 50–900 nm. The UV-Vis spectra of vitamin B12 methanol permeate solutions obtained over seven consecutive days were measured to obtain the corresponding spectra for each day.

In a 7-day dye removal study, high-performance liquid chromatography (HPLC) validated the quantitative results of Vitamin B12 determination based on UV-Vis spectroscopy. Permeate samples were analysed by HPLC over seven days. The concentrations C_f and C_p , determined by high-performance liquid chromatography (HPLC), are used to calculate the rejection rate. The daily rejection rate reported in the HPLC results table is obtained using the following formula:

$$R = \left(1 - \frac{C_p}{C_f}\right) \times 100\% \quad (1)$$

4 Results and Discussion

4.1 Pressure-Permeance Relationship for Methanol Transport

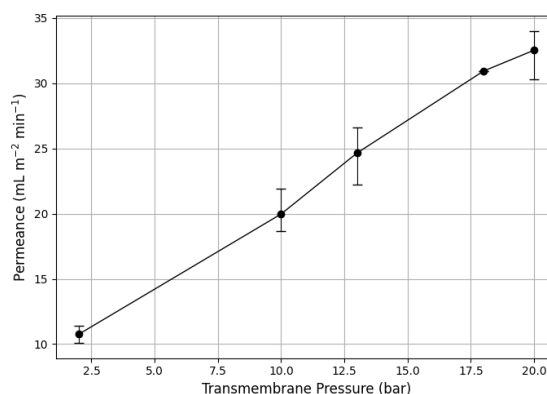


Figure 5: Membrane Permeance as a Function of Transmembrane Pressure

Figure 5 shows that permeance increases consistently with pressure, rising from 10.76 mL m⁻² min⁻¹ at 2 bar to 32.52 mL m⁻² min⁻¹ at 20 bar.

The near-linear trend indicates that methanol transport remains in a regime dominated by viscous flow, where flux is directly proportional to the applied driving force, following Darcy's Law. No evidence of membrane compaction was observed within this pressure range such as the slope decreasing or a plateau in permeance at high pressures. Instead, the membrane structure appears mechanically stable under all tested conditions. The small error bars reflect good reproducibility.

At lower pressures (2–10 bar), permeance increases at a moderate rate, consistent with transport through a dense polymer matrix with limited free volume. Above 10 bar, a slightly steeper increase is observed, which may be attributed to enhanced solvent uptake and swelling of the Matrimid-based matrix, increasing the free volume available for methanol transport. However, minimal deviation from linear behaviour suggests that any swelling effects are slight and that the membrane retains structural integrity throughout operation.

Overall, the pressure–permeance trend indicates that the membrane delivers stable, predictable methanol transport under the tested conditions, supporting the use of these operating pressures in later permeation and selectivity experiments.

4.2 Cage Loading-Permeance Relationship

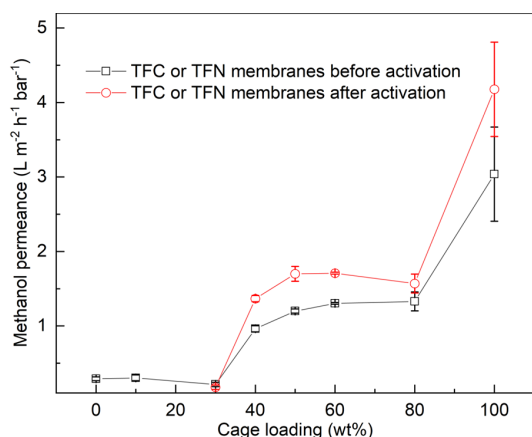


Figure 6: Methanol permeance as a function of cage loading

Figure 6 shows the impact of cage loading on methanol permeance, measured before and after activation. At low cage contents (0–20 wt%), both membrane types exhibit low permeance, indicating that the selective layer remains dense and the cages do not contribute significantly to methanol transport.

A peak in permeance occurs at 40 wt% cage loading, particularly after activation. Activated membranes show a substantial increase compared to non-activated membranes, showing that activation removes residual solvent and opens the cage microporosity, increasing accessible free volume.

Between 40–80 wt%, permeance rises for both membranes, however, activated membranes show higher permeance across all loading concentrations, confirming that activation improves performance. At 100 wt%, permeance increases sharply,

with activated membranes showing the highest values overall.

The data indicates that both cage loading and activation are crucial in enhancing solvent transport.

4.3 Scanning Electron Microscopy

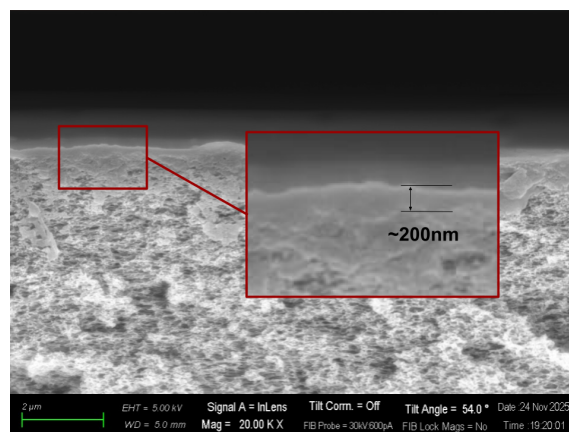


Figure 7: Cross-sectional SEM image showing a dense Matrimid–CPOC-301 selective layer on a porous PAN support, with visible surface irregularities attributed to local cage aggregation.

4.3.1 Cross-Sectional Morphology (SEM)

Figure 7 shows the cross-sectional morphology of the composite membrane, revealing a porous PAN support layer with a thin Matrimid 5218-CPOC-301 selective layer deposited on top. The PAN substrate shows the expected porous structure, while the upper dense layer forms a continuous film across the surface. However, the selective layer is not perfect and contains several raised features or "bumps" that interrupt the morphology. These raised features may arise from local variations in film thickness or cage concentration during casting. Such raised features and local thickness variations have been reported previously as indicators of interfacial defects in mixed-matrix membranes [9]. This highlights the importance of uniform mixing and distribution to prevent inconsistencies in the final membrane.

4.4 Dye Rejection Results

4.4.1 UV-Vis Spectrophotometry

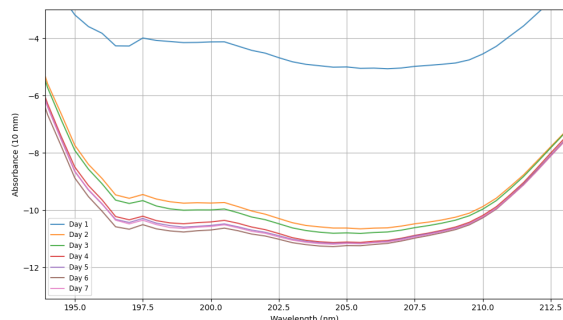


Figure 8: UV-Vis spectra of Vitamin B12 permeate over seven days, showing stabilisation after Day 1 and consistent rejection behaviour.

UV-Vis spectra of the Vitamin B12 methanol permeate samples are shown in Figure 8. Day 1 shows noticeably different absorbance behaviour across the full wavelength range, with a strong peak around 220-240nm and higher readings at shorter wavelengths. From Day 2 onwards, the spectra overlap closely and show a consistent minimum at around 200-205nm. The zoomed-in view shows that absorbance drops sharply between Day 1 and Day 2, after which the curves become much more tightly grouped. One additional anomaly is that Day 7 sits slightly above Day 6, although it still follows the same general shape as the rest of the spectra.

The deviation on Day 1 may be due to bubbles/residue in the cuvette affecting the reading or stabilisation of the membrane. Small variations like the higher absorbance on Day 7 are also expected and can be caused by minor differences in cuvette alignment, cleaning, or handling rather than any chemical change.

4.4.2 High-Performance Liquid Chromatography

Table 1: Vitamin B₁₂ rejection measured by HPLC over seven days.

Day	Rejection (%)
1	86.6
2	100
3	100
4	100
5	100
6	100
7	100

The membrane exhibited a rejection of 86.6 % on Day 1, which is slightly lower than the values observed on following days. This reduced initial re-

jection is explained by early-stage membrane stabilisation. Once stabilised, the membrane demonstrated consistent and complete solution rejection, with 100% Vitamin B12 rejection from Day 2 to Day 7.

This rapid transition to 100% rejection and its stability across multiple days indicate that the membrane structure is intact and defect-free. The results show strong exclusion and suggest that the membrane maintains its selectivity over long-term operation without degradation or performance loss.

4.4.3 Rejection-Molecular Weight Relationship

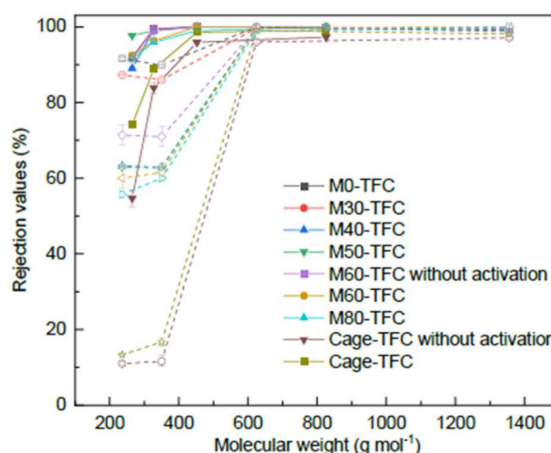


Figure 9: Rejection Rate as a Function of Molecular Weight

Molecular-weight-dependent dye rejection performance of Matrimid-5218-CPOC-301 membranes in terms of Matrimid-5218 fractional composition is shown in Figure 9. Across all membrane types, rejection increases sharply with molecular weight, approaching 100% for solutes above approximately 600-800g/mol, consistent with nano-filtration behaviour. The neutral dyes are shown with a continuous line, whilst the negative dyes are shown with a dashed line.

At lower molecular weights (200-400 g/mol), samples differentiate. The neat polymer membrane (M0-TFC) shows moderate rejection in this range, whereas membranes incorporating CPOC-301 show varying degrees of enhancement depending on cage loading and activation state. Activated membranes generally exhibit higher rejection values than non-activated ones. This is due to removal of residual solvent and opening of microporous cages, resulting in tighter and more selective separation.

4.5 CO₂ Adsorption Isotherms

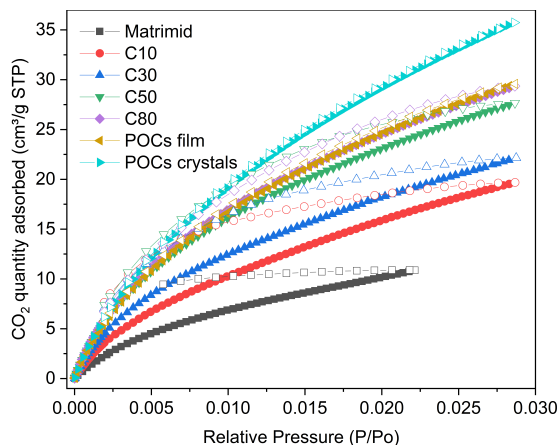


Figure 10: CO₂ adsorption isotherms (273 K) for Matrimid, CPOC-301, and MMMs, showing increased microporosity with higher cage loading.

Figure 10 shows the CO₂ adsorption isotherms (273K) for Matrimid 5218, CPOC-301 and a series of Matrimid/CPOC-301 mixed-matrix composites (C10, C30, C50, C80). The isotherms are in the low pressure region ($P/P_0 = 0.000-0.030$).

All samples show a steep initial rise in CO₂ uptake at very low relative pressures, characteristic of Type I isotherms associated with microporous materials[10]. This indicates strong CO₂ affinity and rapid pore filling within both the polymer matrix and CPOC-301 cages. However, clear differences in adsorption capacity are observed between materials, highlighting variations in porosity.

Matrimid 5218 is a dense polyimide with relatively low surface area, and hence shows the lowest CO₂ uptake across the pressure range. Addition of CPOC-301 into the Matrimid increases adsorption. The composite samples display an increasing CO₂ uptake with increasing CPOC-301 loading, showing that the POC contributes significantly to microporosity.

At higher loadings (C80), the isotherm closely matches the isotherm of the CPOC-301 film, hence the cage molecules retain porosity within the polymer matrix. The CPOC-301 crystalline sample has the highest uptake of CO₂ of all materials, consistent with high packing efficiency and accessibility of the full cage pore network.

Desorption in the CO₂ isotherms show a slight hysteresis, with the desorption branch lying below the adsorption branch for many samples. This indicates slower CO₂ release from narrow micropores, where diffusion is restricted and gas molecules experience stronger confinement.

Overall, the graph confirms that CPOC-301 substantially increases the available adsorption

sites relative to Matrimid 5218, and that the porosity scales with cage loading [11].

4.6 Experimental Challenges

4.6.1 Screen-Coated Membrane Fabrication

Screen coating presented significant challenges in producing uniform Matrimid/CPOC-301 selective layers. Due to the low viscosity of the casting solution and high porosity of the PAN substrate, the liquid film often distributed unevenly across the surface. Small variations in hand motion, glass rod angle and timing created thickness gradients and streaks.

4.6.2 Permeance Testing using Membranes

The inconsistencies resulting from screen coating resulted in variable permeance. Membranes containing thin regions or streaks showed unusually high permeance, as defects created non-selective pathways. On the other hand, areas with a high thickness produced extremely low permeance. This behaviour made it difficult to obtain reproducible flux measurements, and the decision was made to use spin-coated membranes.

4.6.3 Limitations of the SEM Imaging Process

Obtaining high magnification SEM images proved challenging due to the extremely small features of the membrane, which required zooming in to the 10^{-7} m scale to properly show the surface texture and film thickness. Despite sputter coating the samples with gold to improve conductivity, repeated attempts to image at these magnifications resulted in beam-induced burning. Both surface and cross-sectional regions were affected, with the electron beam rapidly damaging the sample before a stable image could be acquired. This is typical for organic polymers which are prone to charging. More advanced methods such as cooling, cryo-SEM, or reducing beam energy may have mitigated this issue.

5 Conclusion

The characterisation and fabrication of Matrimid 5218-CPOC-301 mixed-matrix membranes was successful, and the influence of porous cage incorporation on liquid-phase transport was evaluated. Methanol permeation experiments showed stable, linear pressure-permeance behaviour across 2-20bar, indicating good mechanical integrity and

predictable solvent transport, with an obtained permeance of $32.52 \text{ mL m}^{-2} \text{ min}^{-1} \text{ bar}^{-1}$. Vitamin B12 dye rejection stabilised at 100% from day 2 onwards, confirming that the selective layer remained defect-free during extended operation. CO_2 isotherm analysis further revealed that CPOC-301 increases microporosity to the composite films, with adsorption increasing consistently with cage loading.

SEM imaging highlighted regions of cage aggregation, suggesting that improved dispersion strategies are required to achieve fully uniform

selective layers. Overall, the results show that CPOC-301 can enhance the porosity and performance of Matrimid based membranes, but optimisation of filler distribution remains a key area for future development.

6 Acknowledgements

We would like to sincerely thank Xiang Li for his guidance, expertise, and support during this project.

References

- [1] Bhoga Arundhathi et al. “Advancements in Mixed-Matrix Membranes for Various Separation Applications: State of the Art and Future Prospects”. In: *Membranes (Basel)* 14.11 (2024), p. 224. DOI: 10.3390/membranes14110224.
- [2] Tawsif Siddique, Naba K. Dutta, and Namita Roy Choudhury. “Mixed-Matrix Membrane Fabrication for Water Treatment”. In: *Membranes* 11.8 (2021), p. 557. DOI: 10.3390/membranes11080557.
- [3] MIT News Office. *A new approach could fractionate crude oil using much less energy*. Press release on MIT News. [Online; accessed 2025-12-08]. May 2025. URL: <https://news.mit.edu/2025/new-approach-could-fractionate-crude-oil-using-less-energy-0522>.
- [4] Gui Min Shi et al. “Recent progress of organic solvent nanofiltration membranes”. In: *Progress in Polymer Science* 123 (2021), p. 101470. DOI: 10.1016/j.progpolymsci.2021.101470.
- [5] María de los Ángeles Ramírez-Kantún et al. “Temperature stable, polymeric thin-film composite membrane for hydrogen separation”. In: *Journal of Membrane Science* 683 (2024), p. 122519. DOI: 10.1016/j.memsci.2024.122519.
- [6] Ho Vinh-Thang and Serge Kaliaguine. “Predictive Models for Mixed-Matrix Membrane Performance: A Review”. In: *Chemical Reviews* 113.7 (2013), pp. 4980–5028. DOI: 10.1021/cr3003888.
- [7] Q. Zhang et al. “Mixed-matrix membranes with soluble porous organic molecular cages”. In: *Journal of Membrane Science* 603 (2020), p. 118016.
- [8] Vahid Vatanpour et al. “Polyacrylonitrile in the Preparation of Separation Membranes: A Review”. In: *Industrial and Engineering Chemistry Research* 62.17 (2023), pp. 6537–6558. DOI: 10.1021/acs.iecr.3c00057.
- [9] Biruh Shimekit et al. “Interfacial Defects on Mixed Matrix Membranes and Mitigation Techniques for Gas Separation: A Review”. In: *Applied Mechanics and Materials* 625 (2014), pp. 653–656. DOI: 10.4028/www.scientific.net/AMM.625.653.
- [10] Markus Thommes et al. *Physisorption of gases, with special reference to the evaluation of surface area and pore size distribution (IUPAC Technical Report)*. Tech. rep. IUPAC Technical Report. [Online; accessed 2025-12-08]. International Union of Pure and Applied Chemistry (IUPAC), 2015. URL: https://www.posmat.cefetmg.br/wp-content/uploads/sites/120/2017/11/IUPAC_Recommendations_gas_sorption_2015.pdf.
- [11] M.R. Hasan et al. “How Can the Filler-Polymer Interaction in Mixed Matrix Membranes Improve Membrane Performance?” In: *Membranes* 14.2 (2024), p. 350. DOI: 10.3390/membranes14020350.

Optimisation of the Pd/NF Catalyst for Glycerol Electrochemical Oxidation Reaction

Shuojun Liu and Xinyuan Gu

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

The electrochemical glycerol oxidation reaction (GEOR) provides an appealing approach for turning the excess glycerol into hydrogen and other useful compounds, such as the three-carbon (C3) products. However, it is challenging to achieve a high conversion rate and C3 products selectivity through direct oxidation. For example, CO poisoning can deactivate the catalysts, and severe oxidation can generate C2 and C1 products. Herein, the Pd-based electrocatalysts supported on the Ni foam were designed and produced to improve the GEOR performance. The High-Performance Liquid Chromatography (HPLC) was used to further evaluate the selectivity of the final products under different reaction conditions. The catalyst with a Pd-loading of 0.2 mg/cm² exhibits the optimal performance at the potential of 0.8 V vs. RHE and the reaction duration of 3000 s. Under these optimised conditions, the catalyst delivered the highest selectivity of 48.13% towards glycerate (GLA) and a peak glycerol conversion of 16%, whereas its maximum current density in the CA analysis achieved at 131 mA/cm². Five consecutive electrolysis cycles were performed to estimate the catalyst stability, whereas the selectivity of glycerate and glycerol conversion remaining over 41% and 12%, respectively. These results provide a new insight into the optimal operating conditions for the GEOR by demonstrating the excellent GLA selectivity, glycerol conversion, stability and reusability of the Pd/Ni foam catalysts.

Keywords: Electrochemical glycerol oxidation reaction (GEOR); Pd based electrocatalysts; Glycerate; Selectivity

1. Introduction

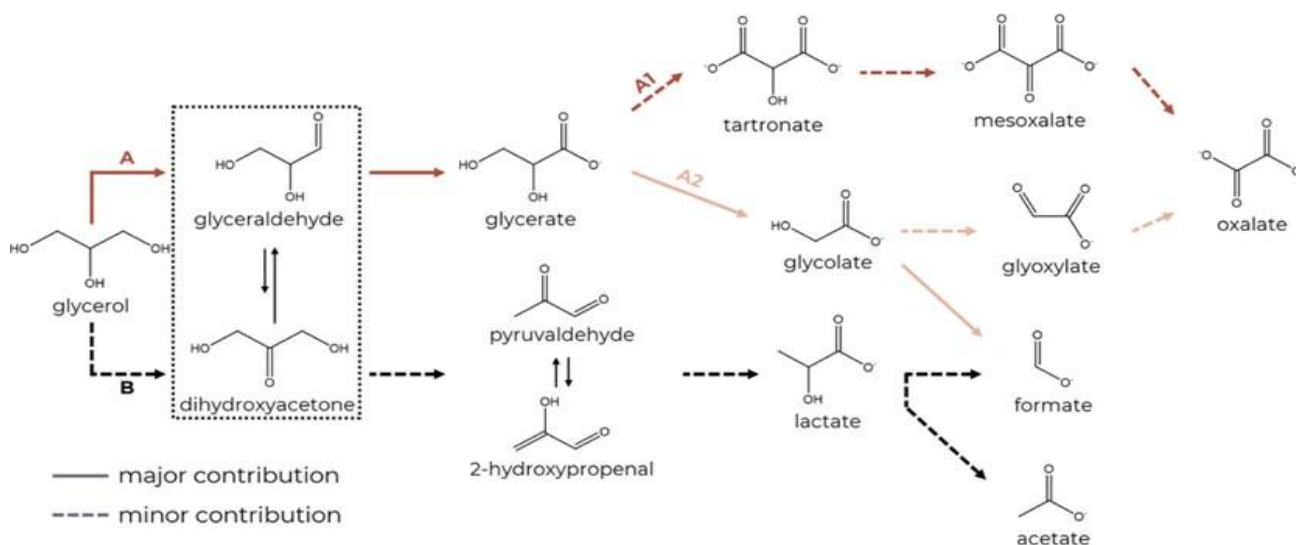


Figure 1: Glycerol oxidation reaction pathway [1].

Biodiesel is an environmentally friendly alternative to petroleum fuels, which is manufactured from vegetable oils, palm oil, and oil-processing waste. [2]. Glycerol (GLY) is the main byproduct of the biodiesel production process, accounting for around 10 wt% of total output, resulting in a global oversupply. This leads to a significant decrease in glycerol prices and influences the sustainability of biodiesel production [3]. Accordingly,

increasing attention has been paid to glycerol as a cheap and abundant feedstock. Moreover, GLY is a highly functional molecule with three hydroxyl functional groups, and it exhibits a high theoretical energy density of 5.0 kWh/kg [4, 5]. Hence, it can serve as a promising platform and be converted into commercially valuable chemicals. The glycerol electrochemical oxidation reaction (GEOR) offers an attractive and sustainable approach to convert

glycerol to value-added compounds, concurrently enabling the coproduction of green hydrogen at the cathode [6]. Compared to conventional water electrolysis, GEOR provides a more energy-efficient route for H₂ generation, as high-value substances rather than low-value oxygen can be generated at the anode. Furthermore, glycerol oxidation with the electrochemical approach takes place under mild conditions that avoid harsh reaction conditions such as high temperature, high pressure, and external oxygen supply [7]. Nevertheless, GEOR proceeds through a complex reaction network that can yield a variety of C1-C3 products such as glycerate (GLA), lactate (LA), glycolate (GA), formate (FA), and so on. Among these, C3 products, particularly GLA, are highly valuable intermediates for pharmaceuticals, biodegradable polymers, and fine chemicals. However, the efficiency and selectivity for these C3 chemicals are difficult to control due to competing pathways and sluggish reaction mechanisms as shown in Figure 1 [8]. Firstly, GLY molecules will be adsorbed on the catalyst surface, then they are converted to form a pair of isomers, glyceraldehyde and dihydroxyacetone intermediates, by dehydrogenation. Subsequent further oxidation and dehydration will take place to generate C3 products such as GLA, LA and TA. Finally, C2 and C1 products can be formed via C-C bond cleavage [6].

As the mechanism of GEOR is highly sluggish and complicated, it is critically necessary to develop and optimize highly efficient and robust electrocatalysts that can enhance the targeted conversion and selectivity of glycerol into valuable C3 products, GLA. Noble metals Au, Pd, and Pt are common metallic catalysts used for GEOR [7]. Pt attracts extensive attention due to its good catalytic activity and stability under both acidic and alkaline conditions [9]. However, it suffers from its sluggish reaction kinetics, and its scarcity significantly raised production costs [10]. The surface oxidation and adsorption of CO can easily occur at low potentials for Pt catalysts, resulting in an inevitable influence on its glycerol oxidation reaction behaviour [8]. Au is more

resistant to deactivation and leaching than Pt, and it prevents the poisoning effect. Nevertheless, some research indicates that catalytic activity of Au is lower compared to that of Pt, and it may and could generate more byproducts from C-C scission. Also, unlike the Pt catalyst, the Au catalyst only works under alkaline conditions [9]. Furthermore, Au typically requires high overpotentials to reach appreciable current densities [11]. In contrast, Pd has emerged as one of the most promising electrocatalysts for GEOR [1]. Pd-based catalysts can reach higher selectivity than Pt at low potential towards GLA under alkaline conditions. This is mainly due to its high catalytic activity towards glycerol oxidation, and the ability to prevent over-oxidation, and Pd has relatively high tolerance to poisoning species [7,10,12].

Nickel foam (NF) was chosen to be used as a support because of its porous structure, offering a large electrochemically active surface area and hence producing a large number of catalytic sites for the oxidation process [13]. In addition, NF has high electrical conductivity, robust mechanical stability, and intrinsic ability to generate abundant active adsorbed hydroxyl species under alkaline conditions that promote the oxidation of adsorbed glycerol species quickly to C3 chemicals and prevent C-C cleavage [10]. Furthermore, Pd-based catalysts supported by NF can achieve high current density that can be attributed to the synergistic effect of Ni and Pd [4]. The alkaline operation conditions are used during the whole experiment, as it can promote dehydrogenation and deprotonation [7,11].

In this study, we mainly focus on the selectivity towards GLA, as it is a valuable raw material that can be used in pharmaceuticals, cosmetics, and fine chemicals. [14] We prepared catalysts with Pd metal overlayers on NF, namely Pd/NF, by using a simple chemical deposition method. The cyclic voltammetry (CV) and linear sweep voltammetry (LSV) were carried out to thoroughly investigate the activity of the catalyst. Comprehensive tests of chronoamperometry (CA) on the Pd/NF catalyst with different Pd loading, applied potential, and reaction

time to optimize the catalyst performance were also performed, aiming to achieve considerable glycerol conversion, high selectivity towards GLA, and reasonable faradaic efficiency. In each experiment, the generated products of GEOR were analysed by high-performance liquid chromatography (HPLC). The conversion, selectivity, and faradaic efficiency were evaluated using the data obtained from CA and HPLC. The results illustrated that the Pd/NF catalyst performs best at a palladium loading of 0.2 mg/cm², operating potential of 0.8 V, and reaction time of 3000 s. Subsequently, 5 consecutive cycle tests were conducted under these optimized conditions to test the cyclicity of the catalyst, and the products for each cycle were analysed via HPLC. Furthermore, the crystal structure of the catalyst was determined by X-ray diffraction (XRD). The study demonstrated that optimizing the amount of deposited Pd on NF and regulating the reaction conditions, such as applied potential and reaction time, significantly enhance the formation of valuable C3 products, especially for GLA, while inhibiting undesired over-oxidation and C–C bond cleavage reaction pathways, the formation of CO that can poison the catalyst is also prevented. The insights gained from this work provide a deeper understanding of the parameters that affect the GEOR activity on Pd-based catalysts and highlight Pd/NF catalyst as an effective platform for the selective electrochemical upgrading of glycerol.

2. Method

2.1 Chemicals and Materials

Palladium (II) chloride (PdCl₂, >99.9%), hydrochloric acid (HCl, 37%), ethanol (C₂H₅OH, >99.5%), and nickel foam were used for making catalysts. To make electrolyte, potassium hydroxide (KOH, >85%), and glycerol (C₃H₈(OH)₃, >99.5%), and deionised water (DI water) were used. For preparation of HPLC vial samples, sulfuric acid (H₂SO₄, 96%) and deionised water were utilised. All chemicals were purchased from Sigma-Aldrich and used as received without further purification.

2.2 Synthesis of Pd/NF Catalyst

The chemical deposition method was carried out to synthesize the Pd/NF catalyst. Firstly, NF pieces (1 cm × 2 cm) were ultrasonically cleaned. The NF was immersed

in ethanol (25 mL) and put in an ultrasonic cleaner for 10 minutes, then discard ethanol and directly immersed it in 1 M HCl (25 mL) for 10 minutes. And repeat these steps for 2 cycles. After that, take the NF out and rinse it with DI water thoroughly to remove residue. Next, for example, to make a Pd/NF catalyst with 0.5 mg/cm² Pd mass loading, 125.3 μL of PdCl₂ solution with concentration of 75 mM was added to 4 mL of DI water in a glass bottle. Subsequently, put the pre-treated NF into the bottle and make sure it was completely immersed in the mixed solution and left it to react at room temperature overnight (about 15 h). After reacting overnight, the solution in the glass bottle turned from yellow to colourless and discard the remaining solution in the bottle. The obtained material was washed with DI water 3-4 times to remove the unreacted residues. Then dry it at 80 °C for 5 hours and denote it as 0.5 mg/cm² Pd/NF, indicating that the Pd mass loading overlaid on the nickel foam was 0.5 mg/cm². Various Pd loadings were obtained by changing the amount of PdCl₂ solution added. The Pd mass loadings and corresponding solution needed are shown in Table 1, and the detailed calculation steps are shown in SI 1.

Table 1: PdCl₂ solution used in catalysts synthesis

Pd Mass Loading / (mg/cm ²)	Volume of PdCl ₂ Solution / μL
0.1	25.1
0.2	50.1
0.5	125.3

2.3 Making Electrolyte

An electrolyte that consists of 3 M KOH and 0.3 M GLY was prepared before electrochemical experiments. The first step was to add 100 g of deionized water to the electrolyte bottle. Then 44.73 g of KOH solid was gradually added to the deionized water half by half to prevent the overheating caused by the exothermic dissolution of KOH. Followed by adding 5.52 g of glycerol (GLY) to the mixture and shake the mixture to ensure chemicals were completely and uniformly mixed. Finally, add the deionized water until the total mass of the electrolyte reached 230 g.

2.4 Electrochemical Experiment

To perform this experiment, a typical 3-electrodes batch system was employed, the NF/Pd catalyst clipped by a platinum sheet electrode holder was used as the working electrode, the graphite rod acted as the counter electrode, and the reference electrode was the Hg/HgO electrode. 10mL of the electrolyte was added to the electrochemical system. These electrodes were all connected to a booster to extend the current range and ensure accurate potential control during high-current-density electrolysis.

The electroanalysis was carried out by using cyclic voltammetry (CV), linear sweep voltammetry (LSV), electrochemical impedance spectroscopy (EIS), and constant voltage chronoamperometry (CA) analysis. All electrochemical experiments were controlled and recorded by implementing NOVA software. During experiments, measured potentials were converted to the reversible hydrogen electrode (RHE) scale according to the following equation:

$$E \text{ (vs. RHE)} = E \text{ (vs. Hg/HgO)} + 0.098 + 0.059 \times \text{pH}$$

2.5 CV, LSV and EIS

The CV test was conducted at a scan rate of 0.1 V/s for 10 cycles to examine the catalytic activity and electrochemical characteristics. The LSV test was performed at a scan rate of 0.01 V/s to verify the reliability of the results obtained from the CV tests and further evaluate the catalytic activity and electrochemical characteristics of the Pd/NF catalysts by determining the onset potential and its peak current density. An EIS test was subsequently undertaken, and the frequency response analyser (FRA) measurement was involved in the EIS test, with a frequency range of 100,000 Hz to 0.02 Hz.

2.6 CA

Prior to the CA measurements, the electrocatalysts were activated through five CV scans. Each CA test was employed using a fresh 10 mL electrolyte solution for ensuring consistent reaction conditions. To further assess the optimal operating condition of electrocatalyst, constant potentials of 0.7 V, 0.8 V and 0.9 V (vs. RHE) and different reaction times of 600 s, 1200 s, 1800 s, 2400 s, and 3000 s were applied. Furthermore, the CA test was repeated for five consecutive cycles to assess the durability and stability of the Pd/Ni foam catalyst.

2.7 Catalyst Characterization

The crystalline structures of catalysts were evaluated by XRD patterns. A PANalytical X'Pert Pro Multi Purpose Diffractometer using Cu K α radiation with K α =1.5186 Å was used to investigate the XRD patterns. Patterns were recorded for 2 θ values ranging from 20.0° to 90.0°

2.8 Product Analysis—HPLC

The electrolyte was collected after CA experiments and analysed by using an Agilent Technologies 1260 Infinity II HPLC equipped with a Coregel 87H3 column and a UV detector and a refractive index detector (RID). During testing, the wavelength for the UV detector is 210 nm and the temperature of column and RID is 40 °C and 35 °C respectively. The effluent used was 0.005 M H₂SO₄ with a flow rate of 0.4 mL/min. The sample consisted of 400 μ L H₂SO₄, 200 μ L deionized water, and 100 μ L electrolyte, so the electrolyte was diluted 7 times. 1 μ L of sample was injected each time for analysis. The concentration of each chemical of the sample in mmol/l could be obtained through qualitative identification based on retention time and quantitative analysis by integrating the peak area and comparison with calibration standards.

The faradaic efficiency (FE), selectivity, and conversion could be computed based on data obtained from HPLC.

The FE of products was computed using the number of electrons transferred and the total charges passed through the circuits, it illustrates the selectivity of the catalyst towards the products, and the formula is shown below.

$$FE = \frac{n_e \times m_{product} \times F}{Q} \times Dilution Ratio \times 100\%$$

Where n_e is the number of electrons transferred, $m_{product}$ is moles of products, F is faraday constant (96485 C/mol), Q is the total charge passed, and *Dilution Ratio* is 7.

The selectivity was obtained based on moles of the product formed and moles of reactant reacts.

$$Selectivity = \frac{n_{product} \times CN_{product}}{(n_0 - n_e) \times CN_{reactant}} \times 100\%$$

Where $n_{product}$ is moles of products formed, $(n_0 - n)$ indicates the difference of the reactant in the electrolyte before and after the reaction, and CN is the carbon number.

The equation of glycerol conversion is shown following:

$$X_{GLY} = \frac{n_{GLY,0} - n_{GLY,e}}{n_{GLY,0}} \times 100\%$$

Where $n_{GLY,0}$ is the molar amount of GLY in electrolyte

before reacting and $n_{GLY,e}$ is the remaining moles of GLY in electrolyte after reaction. The half reactions of each product listing the number of electrons transferred could be found in SI 2.

3. Results and Discussion

3.1 Various Pd Loading

3.1.1 Structural Characterization of Catalysts

The crystalline structure of Pd/NF catalyst was evaluated by the XRD patterns. The XRD patterns of catalysts with various Pd loadings and reference peaks for Ni and Pd are shown in Figure 2. All catalysts show 3 intensive peaks that were aligned well with face-centred cubic (fcc) Ni (JCPDS 04-0850). There is no peak shift between the characteristic reflections of pure Ni and pure Pd, indicating that Ni and Pd are not alloyed. Instead, Pd is overlaid or highly dispersed on the NF.

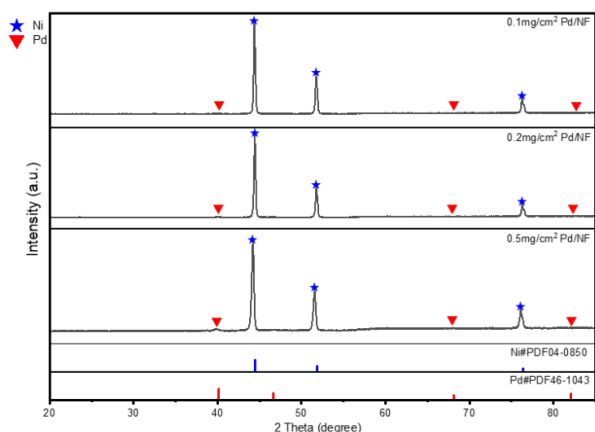


Figure 2: XRD patterns for various Pd loading catalysts.

From Figure 2, as Pd mass loading increases, a weak peak appears at $2\theta \approx 40^\circ$, which corresponds to the (111) reflection of metallic fcc Pd (JCPDS 00 046-1043). This indicates the Pd crystallite size gets larger with increasing Pd loading, demonstrating that Pd nanoparticles are deposited on the Ni foam rather than forming a Pd–Ni alloyed phase.

3.1.2 Catalytic Activity

The electrocatalytic performances of Pd/NF catalysts with different Pd loading were examined in an alkaline electrolyte that consisted of 3 M KOH and 0.3 M GLY with a typical three electrode batch system. Figure 3a is the cyclic voltammograms (CVs) for Pd/NF catalysts with various Pd mass loading. It can be observed that CV

involves 2 scans, which are the forward scan and the backward scan. The anodic peak during forward scan indicates palladium oxidation, the cathodic peak in backward scan represents reduction of PdO_x [18]. All Pd/NF catalysts have different onset potentials for GEOR to begin, which are around $E_{onset}=0.6V, 0.65 V, 0.7 V$ for $0.5 \text{ mg/cm}^2, 0.2 \text{ mg/cm}^2, 0.1 \text{ mg/cm}^2$ Pd loading respectively, and reached the highest peak in the forward scan at $0.93 V$ vs. RHE. The trends of CVs illustrate that GEOR catalytic activity is dependent on Pd mass loading. As Pd loadings of catalysts increase, the peak intensity increases. The current density of the catalyst with Pd loading of 0.5 mg/cm^2 peaked at around 256 mA/cm^2 , which was almost 2.5 times as much as 0.1 mg/cm^2 Pd/NF. 0.2 mg/cm^2 Pd/NF exhibited slightly lower current density than that of 0.5 mg/cm^2 Pd/NF, which peaked at near 174 mA/cm^2 . Higher peak indicates higher catalytic activity of the catalyst. This can be attributed to that more active site with increasing Pd loading, hence higher electron transfer and better reaction kinetics. Electrochemical impedance spectroscopy (EIS) was carried out at $0.8 V$ vs RHE with the frequency ranging from 10000Hz to 0.02Hz to evaluate the kinetics of Pd/NF catalysts as shown in Figure 3c. The EIS were modelled by using an equivalent circuit consisting of a solution resistance (R_s) in series with a parallel combination of the charge-transfer resistance (R_{ct}) and a constant phase element (CPE). Where R_s is the ohmic resistance of the electrolyte and cable, R_{ct} reflects the interfacial kinetics of glycerol oxidation. Figure 3c illustrates that 0.5 mg/cm^2 Pd/NF catalyst exhibited the smallest value of R_{ct} [15], which demonstrates that it has efficient electron-transfer capability across the electrode/electrolyte interface during the glycerol oxidation reaction. Consequently, 0.5 mg/cm^2 Pd/NF catalyst has the highest catalytic kinetics. The catalyst with Pd loading of 0.2 mg/cm^2 displays moderate radius, hence the reasonable magnitude of R_{ct} and catalytic kinetics. All catalysts show the same x-interception, suggesting that the value of R_s remains constant. Hence improved performance as the result of increased catalytic activity at the Pd/NF interface rather than changes in electrolyte resistance. The result of EIS is consistent with CV, where 0.1 mg/cm^2 Pd/NF shows the poorest catalytic activity.

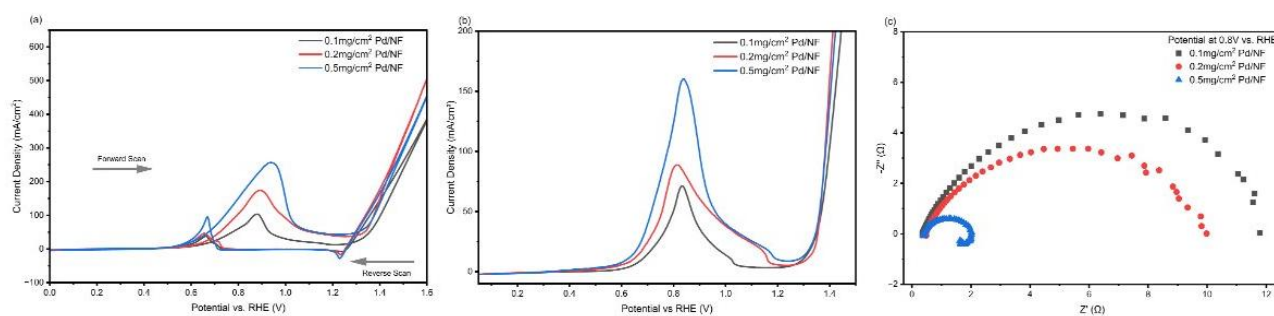


Figure 3: (a) CV curves for various Pd loading catalysts in 3M KOH and 0.3M GLY electrolyte at a scan rate of 10 mV/s. (b) LSV curves for various Pd loading catalysts in 3M KOH and 0.3M GLY electrolyte after IR correction. (c) Electrochemical impedance spectroscopy (EIS) Nyquist plots for various Pd loading catalysts in 3M KOH and 0.3M GLY electrolyte at a potential of 0.8V vs. RHE.

3.1.3 Selectivity & Conversion

The chronoamperometry (CA) tests of GEOR using Pd/NF catalysts were conducted at 0.8V for 2400 s. Then electrolytes before and after reaction were collected for HPLC analysis. As can be seen from Figure 4a, the conversion increases with the Pd mass loading, which is ascribed to increasing the number of active sites as the Pd loading increases and higher activity for higher mass loading, that is observed in Figure 3a. The catalyst with 0.5 mg/cm² Pd loading has the highest conversion of 15.3%, and 0.2 mg/cm² Pd/NF catalyst has a similar value of conversion of 14.7%. The reason for the similar conversion value is the mass transport limitation. It may occur at higher mass loading even though the current density is higher. However, selectivity does not follow a similar tendency as the conversion, the highest selectivity to GLA of 43.91% was achieved for the 0.2 mg/cm² Pd/NF catalyst. This may be mainly because the uniform dispersion of Pd on the NF for this catalyst, contributing to higher fraction of active sites that favour the selective partial oxidation pathway to GLA. Also, uniformly distribution of Pd/NF in the catalyst can lead glycerol molecules to be completely oxidised to carbonates. In addition, the catalyst may induce the optimised balance between active Pd sites and NiOOH, an active oxidising species that will promote further oxidation.

For 0.1 mg/cm² Pd/NF catalyst, the selectivity towards GLA is the lowest, while that towards FA and OA is obviously higher than the other catalysts. The catalyst surface is enriched with Ni content. The rich Ni content leads to the more sites for the formation of Ni(OH)_x and then can be further oxidised to NiOOH, which is an active

oxidising species that can provide reactive oxygen species and facilitate the adsorption of hydroxyl species from the GLY and promote the oxidation of intermediates, hence enhancing C-C bond cleavage.[16,17] High Ni content promotes partial C-C bond cleavage and high overpotentials for GLY oxidation is needed, leading to the selectivity towards OA and GA is high. Moreover, less Pd sites are exposed, resulting in insufficient active sites for the oxidation of GLY to take place, resulting in reducing selectivity. Whereas for 0.5 mg/cm² Pd/NF catalyst, it has a Pd-rich surface, it has faster reaction kinetics that may keep intermediates on the surface long enough to be further oxidised and promote C-C cleavage. Moreover, as indicated by XRD, the 0.5 mg/cm² Pd/NF catalyst has larger crystalline structure which may lead to a reduction in surface area and hence the catalytic activity. In addition, the active site for 0.5 mg/cm² Pd/NF is too bulky, resulting in mass transfer limitation, the molecules cannot diffuse to the active sites for the reaction to occur, reducing effective utilisation of active sites. Thus, the catalyst with 0.2 mg/cm² Pd mass loading was used for further investigation.

3.2 Various Potential Applied

Figure 4b presents the performance of catalyst with a Pd loading of 0.2 mg/cm² under various potential applied. As the potential increases from 0.7 V to 0.8 V vs. RHE, both the glycerol conversion and selectivity towards glycerate significantly improves, rises from 6.97% and 22.18% to 14.73% and 43.91% respectively, indicating that the glycerol oxidation is more favoured at higher potentials [1]. However, when the potential further rises to 0.9V vs. RHE, the glycerol conversion and GLA selectivity show a

reducing trend to 4.79% and 34.89%, denoting that 0.8V vs. RHE is the optimal potential for achieving the greatest catalytic performance. Figure 4c illustrates the CA responses of the 0.2mg/cm² Pd/NF at 0.9V vs. RHE over 2400s, the current density is initially very high and followed by a sharp and rapid decay. For each potential pulse applied for every 300s, its steady state current density achieves at approximately 10 mA/cm². This phenomenon is attributed to the excessive oxidation of Pd-OH to PdO_x at high potentials, which leads to catalyst deactivation. The formation of PdO_x layer covers the Pd surface, lowering its affinity for glycerol adsorption, thereby weakening the interaction between glycerol molecules and active sites of Pd. This oxidation layer blocks the Pd active sites, the catalytic surface becomes limited, which is also a major restriction for further glycerol oxidation reaction. Thus, both conversion and the selectivity towards the C3 products are reduced [18].

3.3 Various Reaction Time

To optimise the reaction conditions for the catalyst, the

time-dependent GEOR performance at a potential of 0.8V vs. RHE is presented by Figure 4d. The glycerol conversion and the selectivity toward C3 products, especially glycerate, exhibit a steady upward trend as time increases, obtaining a value of 27.68% at 600 s to 48.13% at 3000 s for selectivity and an increasing value from 5.35% at 3000 s for selectivity and an increasing value from 5.35% at 600 s to 16% at 3000 s for glycerol conversion. According to previous studies [19], the applied potential of 0.8V vs. RHE represents a mild oxidation condition, whereas the glycerol electrochemical oxidation reaction primarily follows a partial oxidation pathway to form C3 products. In contrast, the formation of C1 and C2 species involves deeper oxidation and the C-C bond cleavage, which requires a stronger oxidation condition and a higher density of the Pd active sites. Consequently, as the reaction time extends, more glycerol is converted into C3 products, leading to a steadily increasing C3 product selectivity. Meanwhile, the selectivity of C1 and C2 products only fluctuates slightly under moderate potential conditions, due to the insufficient oxidation conditions for C-C bond breaking.

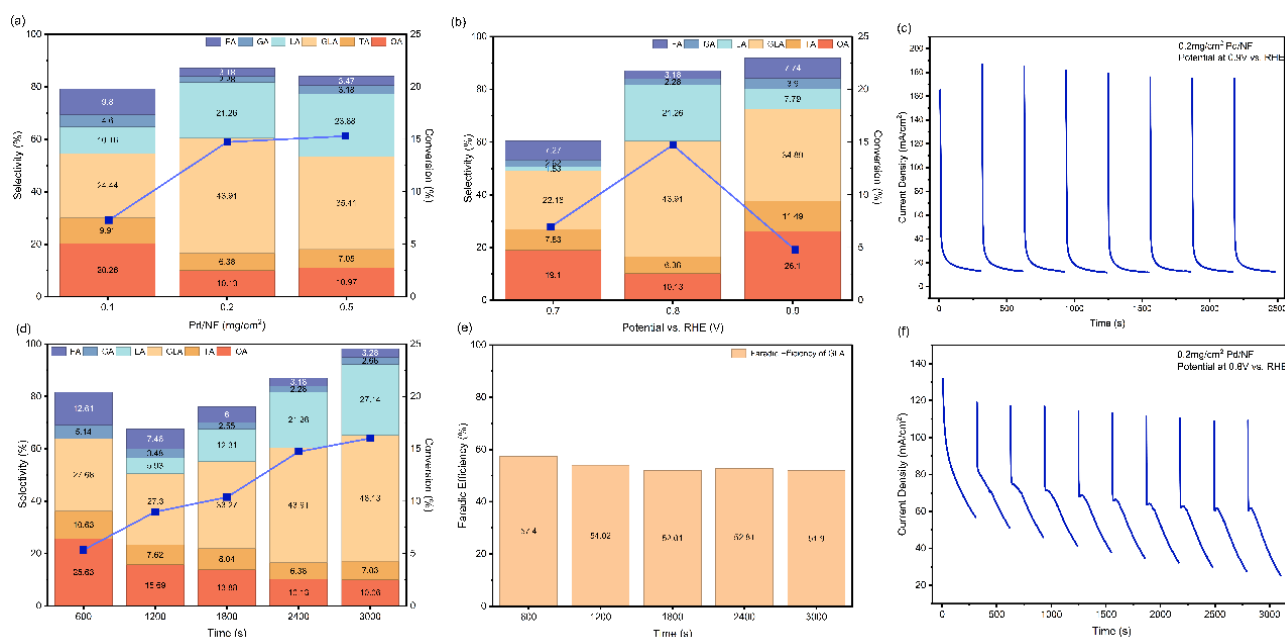


Figure 4: (a) Selectivity and glycerol conversion for various Pd loading catalysts at a potential of 0.8 V vs. RHE for 2400s. (b) Selectivity and glycerol conversion for 0.2mg/cm² Pd/NF at various potential for 2400s. (c) Chronoamperometry (CA) analysis at 0.9 V vs. RHE for 0.2mg/cm² Pd/NF during 2400s with potential pulse for every 300s. (d) Selectivity and glycerol conversion for 0.2mg/cm² Pd/NF at 0.8V vs. RHE for various electrolysis durations. (e) Faradaic efficiency of glycerate at a potential of 0.8 V vs. RHE for various electrolysis durations using 0.2mg/cm² Pd/NF. (f) CA analysis at 0.8 V vs. RHE for 0.2mg/cm² Pd/NF during 3000s with potential pulse for every 300s.

As shown in Figure 4e, the faradaic efficiency of GLA remains relatively stable, deviates between 51.9% to

57.4%, confirming the efficient electron utilisation for the GLA generation. However, it is worth noting that the

faradaic efficiency of GLA may not exclusively reflect to the electrochemical oxidation of GLA. During the dehydrogenation of glycerol, intermediates such as glyceraldehyde (GLAD) and dihydroxyacetone (DHA) formed, which can experience further oxidation through chemical pathways to generate glycerate with hydroxide ions. Therefore, the final GLA production includes both electrochemical and chemical contribution. Since the yield of glycerate is raised, it results in an overestimated faradaic efficiency level. [20]

Additionally, the CA graph given in Figure 4f reveals the periodic potential pulses for 3000s. A sharp current density at the start of the potential pulse is observed and subsequently shows a decreasing trend. As the glycerol oxidation processes, the oxidation layer PdO_x forms and covers on the catalyst surface, blocking the active sites for glycerol adsorption, which affects the activity of the catalysts [21]. Although the presence of PdO_x influences the catalyst activity, the current densities at later cycles still remain over 25 mA/cm^2 , confirming that the 0.2 mg/cm^2 Pd/NF catalyst maintains an excellent stability

and activity under this moderate potential.

3.4 Cyclic Tests

Figures 5a and 5b clearly demonstrate that the 0.2 mg/cm^2 Pd/NF catalyst exhibits an excellent cyclic durability and structural stability at 0.8 V vs RHE. The glycerol conversion remains nearly constant over the five consecutive cycles, and the selectivity towards the C3 products, particularly glycerate stays consistently high, suggesting that the reaction pathway predominantly governed by partial oxidation rather than deeper oxidation. Meanwhile, the XRD profile of the catalyst before and after reaction manifests the identical diffraction peaks corresponding to the Pd and Ni shown in Figure 5b, verifying that there is no significant Pd transformation or structural reconstruction occurs during the GEOR cycles. Additionally, the stable selectivity also implies that the Pd active sites is largely preserved, preventing the catalyst deactivation. Overall, these experimental results validate that the 0.2 mg/cm^2 Pd/NF catalyst exhibits an excellent robustness and reusability during the GEOR processes.

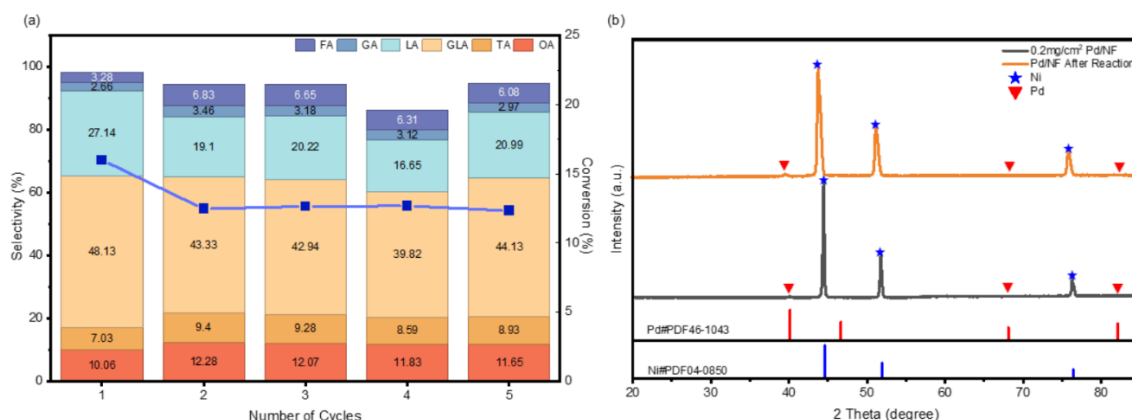


Figure 5: (a) Selectivity and glycerol conversion for 0.2 mg/cm^2 Pd/NF for 5 consecutive cycles at the potential of 0.8 V vs. RHE, each cycle lasts for 3000s. (b) XRD patterns for 0.2 mg/cm^2 Pd/NF before and after reaction.

4. Conclusion

In this study, a Pd/NF catalyst was systematically investigated for glycerol electrochemical oxidation reaction (GEOR) in the alkaline electrolyte. By synthesizing the tests on the catalysts with different Pd loading, 0.2 mg/cm^2 is considered to be the optimal Pd loading since it illustrates the highest glycerate selectivity (48.13%). To enhance the reaction condition for higher selectivity, the applied potential

was varied to identify the optimal value. It was found that at a moderate potential of 0.8 V vs RHE delivers the best performance on both the glycerol conversion and the selectivity toward GLA. Time-dependent experiments revealed that both glycerol conversion and C3 product selectivity increased steadily as reaction time increases, while the C1 and C2 products changed only slightly, supporting the fact that the reaction predominantly follows a partial oxidation pathway rather than deeper oxidation and C–C bond

cleavage. The CA profiles delivered stable current density responses, indicating the strong electrochemical durability of the catalysts. Furthermore, cyclic tests and XRD diffraction patterns verified that the Pd/NF catalyst retained its structural integrity and Pd active phase after 5 cycles. Overall, the results discussed above highlight the remarkable robustness, stability, reusability and its high GLA selectivity of the Pd/NF catalyst, demonstrating its promising Pd loading and potential applied for efficient glycerol electrochemical oxidation reaction towards the valued C3 chemicals.

5. Outlook

Although this study demonstrates that Pd-coated nickel foam is an effective and stable electrocatalyst for the selective electrochemical oxidation of glycerol towards C3 products, further optimisation is still required to improve the performance of the catalyst. Future work will focus on other factors that may affect catalytic activity such as temperature. Moreover, as we mentioned above, the conversion and selectivity of GLA and LA increase with reaction time. Thus, the effect of longer reaction time for CA test on selectivity and conversion will be investigated. In addition, the performance of using Pd-M alloys (M=Au, Pt, Bi, etc.) as electrocatalysts can be explored and compared to that of NF/Pd catalyst. Furthermore, techniques such as scanning electron microscopy (SEM) can be used to analyse the morphology of the catalyst. X-ray photoelectron spectroscopy (XPS) can be applied to better understand the surface composition and chemical state of the catalyst; hence a reasonable mechanism can be deduced. Finally, to improve the sustainability and economic viability of GEOR, a renewable electricity power source can be considered to use for GEOR.

6. Acknowledgements

We want to express our appreciation to Yabo Qi for all of his help, patience, and continuous support throughout this project. His insightful suggestions and technical assistance were crucial to the accomplishment of this project.

7. Reference:

1. Terekhina, I., White, J.B., Cornell, A. and Johnsson, M. (2023). Electrocatalytic Oxidation of Glycerol to Value-Added Compounds on Pd Nanocrystals. *6*(13), pp.11211–11220. doi:https://doi.org/10.1021/acsnm.3c01236.
2. Villa, A., Dimitratos, N., Chan-Thaw, C.E., Hammond, C., Prati, L. and Hutchings, G.J. (2015). Glycerol Oxidation Using Gold-Containing Catalysts. *Accounts of Chemical Research*, *48*(5), pp.1403–1412. doi:https://doi.org/10.1021/ar500426g.
3. Yang, F., Hanna, M.A. and Sun, R. (2012). Value-added Uses for Crude glycerol—a Byproduct of Biodiesel Production. *Biotechnology for Biofuels*, [online] *5*(1), p.13. doi:https://doi.org/10.1186/1754-6834-5-13.
4. Fan, L., Liu, B., Liu, X., Senthilkumar, N., Wang, G. and Wen, Z. (2020). Recent Progress in Electrocatalytic Glycerol Oxidation. *Energy Technology*, *9*(2), p.2000804. doi:https://doi.org/10.1002/ente.202000804.
5. Simões, M.R., Stève Baranton and Christophe Coutanceau (2010). Electro-oxidation of glycerol at Pd based nano-catalysts for an application in alkaline fuel cells for chemicals and energy cogeneration. *Applied Catalysis B-environmental*, *93*(3-4), pp.354–362. doi:https://doi.org/10.1016/j.apcatb.2009.10.008.
6. Yu, X., Santos, White, J.B., Germán Salazar-Alvarez, Lars, Cornell, A. and Johnsson, M. (2021). Electrocatalytic Glycerol Oxidation with Concurrent Hydrogen Evolution Utilizing an Efficient MoOx/Pt Catalyst. *Small*, *17*(44). doi:https://doi.org/10.1002/smll.202104288.
7. Li, T. and Harrington, D.A. (2021). An Overview of Glycerol Electrooxidation Mechanisms on Pt, Pd and Au. *ChemSusChem*, *14*(6), pp.1472–1495. doi:https://doi.org/10.1002/cssc.202002669.
8. Hu, X., Jun Lü, Liu, Y., Chen, L., Zhang, X. and Wang, H. (2023). Sustainable catalytic oxidation of glycerol: a review. *Environmental Chemistry Letters*, *21*(5), pp.2825–2861. doi:https://doi.org/10.1007/s10311-023-01608-z.
9. Ashutosh Namdeo, S.M. Mahajani and Suresh, A.K. (2016). Palladium catalysed oxidation of glycerol—Effect of catalyst support. *Journal of Molecular Catalysis A Chemical*, *421*, pp.45–56. doi:https://doi.org/10.1016/j.molcata.2016.05.008.
10. Zhang, Y., Wang, L., Pan, S., Zhou, L., Zhang, M., Yang, Y. and Cai, W. (2024). Improving the Electrochemical Glycerol-to-Glycerate Conversion at Pd Sites via the Interfacial Hydroxyl Immigrated from Ni Sites. *Molecules*, [online] *29*(16), p.3890. doi:https://doi.org/10.3390/molecules29163890.

11. Antolini, E. (2019). Glycerol Electro-Oxidation in Alkaline Media and Alkaline Direct Glycerol Fuel Cells. *Catalysts*, 9(12), p.980. doi:<https://doi.org/10.3390/catal9120980>.
12. Bianchi, C.L., Canton, P., Dimitratos, N., Porta, F. and Prati, L. (2005). Selective oxidation of glycerol with oxygen using mono and bimetallic catalysts based on Au, Pd and Pt metals. *Catalysis Today*, 102-103, pp.203–212. doi:<https://doi.org/10.1016/j.cattod.2005.02.003>.
13. Vij, V., Sultan, S., Harzandi, A.M., Meena, A., Tiwari, J.N., Lee, W.-G., Yoon, T. and Kim, K.S. (2017). Nickel-Based Electrocatalysts for Energy-Related Applications: Oxygen Reduction, Oxygen Evolution, and Hydrogen Evolution Reactions. *ACS Catalysis*, 7(10), pp.7196–7225. doi:<https://doi.org/10.1021/acscatal.7b01800>.
14. Habe, H., Fukuoka, T., Kitamoto, D. and Sakaki, K. (2009). Biotechnological production of d-glyceric acid and its application. *Applied Microbiology and Biotechnology*, 84(3), pp.445–452. doi:<https://doi.org/10.1007/s00253-009-2124-3>.
15. Si, D., Xiong, B., Chen, L. and Shi, J. (2021). Highly selective and efficient electrocatalytic synthesis of glycolic acid in coupling with hydrogen evolution. *Chem Catalysis*, 1(4), pp.941–955. doi:<https://doi.org/10.1016/j.checat.2021.08.001>.
16. Zhang, Q., Zhang, X., Liu, B., Jing, P., Xu, X., Hao, H., Gao, R. and Zhang, J. (2025). Active Hydroxyl-Mediated Preferential Cleavage of Carbon-Carbon Bonds in Electrocatalytic Glycerol Oxidation. *Angewandte Chemie International Edition*, 64(9), pp.e202420942–e202420942. doi:<https://doi.org/10.1002/anie.202420942>.
17. Zahra Hagheh Kavousi, Badie, C., Vyshnav Kannampalli, Massomeh Ghorbanloo, wang, Q., Hazar Guesmi, Santinacci, L., Yaovi Holade and Mikhael Bechelany (2025). Atomically bridged palladium between nickel species and carbon microfibers and the subsequent conversion into free-standing and electrocatalytically active multifunctional electrodes. *Journal of Materials Chemistry A*, [online] 13(40), pp.34231–34248. doi:<https://doi.org/10.1039/d5ta02374b>.
18. Xia, T., Yang, J., Ren, Q., Fu, Y., Zhang, Z., Li, Z., Shao, M. and Duan, X. (2024). Promoting Alcohols Electrooxidation Coupled with Hydrogen Production via Asymmetric Pulse Potential Strategy. *Angewandte Chemie International Edition*, 64(9). doi:<https://doi.org/10.1002/anie.202420992>.
19. Zhou, C., Chen, J., Zhao, J., Meng, Y., Li, Z., Meng, X. and Wang, J. (2024). Glycerol Electrooxidation to Value-Added C₁–C₃ Chemicals: Mechanism Analyses, Influencing Factors, Catalytic Regulation, and Paired Valorization. *Renewables*, 2(2), pp.89–110. doi:<https://doi.org/10.31635/renewables.024.202400052>.
20. Li, J., Tu, Y., He, K., Chen, C., Liang, L., Ruan, C. and Zhang, Q. (2025). Mechanistic Insights into Glycerol Oxidation to High-Value Chemicals via Metal-Based Catalysts. *Molecules*, [online] 30(6), p.1310. doi:<https://doi.org/10.3390/molecules30061310>.
21. Cruz Ribeiro, J.Y., Santos da Silva, R., Salazar-Banda, G.R. and Barrios Eguiluz, K.I. (2023). Pd-Ni-B/C Nanocatalysts for Electrochemical Oxidation of Ethanol in Alkaline Media. *Catalysis Research*, [online] 03(01), pp.1–19. doi:<https://doi.org/10.21926/cr.2301005>.

Characterising SAF: Evaluating the Physicochemical Properties of Sulphonated Alkyl Furoates

Faith Budge and Loris Butler

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

To address the environmental burden of conventional surfactants, there is a growing need for renewable alternatives. This study characterises sulphonated alkyl furoates (SAF), a proposed bio-derived anionic surfactant, to assess its suitability as a detergent. The physicochemical properties of SAF were evaluated through hydrophilic-lipophilic balance (HLB) determination, dynamic wetting analysis, and phase behaviour mapping, using sodium dodecyl sulphate (SDS) and sodium dodecyl benzene sulphonate (SDBS) as benchmarks. Through iterative emulsion screening using 1-dodecanol and rapeseed oil, the experimental HLB of SAF-12 was determined to be 40.7, indicating a high degree of hydrophilicity. In sessile drop contact angle analysis, SAF-12 demonstrated wetting performance comparable to SDBS, effectively matching the solubility of SDS and the detergency of SDBS. Additionally, rheological profiling and polarised light microscopy (PLM) identified distinct micellar, hexagonal, and lamellar structures, allowing for the construction of a binary phase diagram. These results suggest that SAF offers a viable, high-performance renewable alternative for industrial formulation.

Keywords: Hydrophilic-Lipophilic Balance (HLB), Phase Behaviour, Polarised Light Microscopy (PLM), Surfactants, Wetting Dynamics

1 Introduction

The surfactant industry is experiencing a substantial transition as manufacturers seek to replace conventional petrochemical derived molecules with renewable, environmentally compatible alternatives. Anionic surfactants remain a significant presence in household and industrial formulations, contributing between 15 - 40% of the typical detergent formulation (Scheibel, 2004). Out of these surfactants, linear alkylbenzene sulphonates (LAS) account for more than 40% of global surfactant production (Kaida et al., 2021). LAS is characterised by a sulphonate head group linked to an alkylbenzene chain that provides hydrophobicity. Although this architecture provides strong detergency, the aromatic ring structure is relatively resistant to microbial degradation, resulting in its persistence through wastewater treatment processes and raising concerns regarding potential ecological effects (Kaida et al., 2021). These drawbacks have accelerated interest in renewable alternatives capable of matching the performance of LAS while reducing environmen-

tal burdens.

Sulphonated Alkyl Furoates (SAF) have recently been proposed as a promising bio-derived anionic surfactant that may satisfy these criteria. SAF is produced by esterifying biomass derived furoic acid with renewable fatty alcohols, followed by sulphonation of the furoate ester (Al Ghatta et al., 2022). The resulting structure, illustrated in Figure 1, mirrors that of LAS but replaces the non-polar benzene ring with a furan ring.

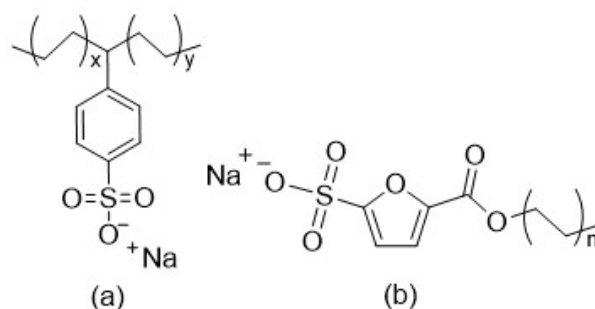


Figure 1: Generalised chemical structures of (a) linear alkylbenzene sulphonate (LAS) and (b) sulphonated alkyl furoate (SAF).

This modification introduces a heteroatom into the hydrophobic region of the molecule, altering

the electronic environment of the ring: the oxygen atom in furan contributes both polarity and hydrogen bonding capability, which are absent from the aromatic core of LAS. Existing reports indicate that this increased polarity may enhance hard water tolerance and yield favourable critical micelle concentrations (CMC) (Al Ghatta et al., 2022). However, the implications of these structural differences for the behaviour of SAF in bulk solution, particularly its phase structure and interfacial properties, remain insufficiently understood.

Although initial studies have described SAF synthesis and basic solubility, comprehensive physicochemical characterisation is still lacking. By contrast, systems based on LAS have been extensively studied, with well established relationships between concentration, structure, and viscosity. Surfactant solutions typically pass through predictable structural regimes, such as spherical micelles, wormlike micelles and ultimately hexagonal or lamellar liquid crystalline phases as their concentration increases. These phase transitions have a direct influence on the rheology and behaviour of formulated products. Currently, however, no detailed phase diagram exists for the SAF water system, limiting the ability to formulate products containing this emerging bio based surfactant.

Another obstacle to the industrial implementation of SAF arises from the lack of appropriate theoretical tools for predicting its emulsification characteristics. The hydrophilic-lipophilic balance (HLB) is widely employed to estimate the suitability of a surfactant for oil-in-water or water-in-oil emulsions. Approaches such as the Davies group contribution method assign numerical values to individual functional groups (Davies, 1957). These approaches do not include a term for the sulphonated furan ring present in SAF, meaning that its HLB cannot be reliably computed using standard models and must instead be obtained experimentally.

The cleaning performance of SAF ultimately depends on more than its solubility and structure; effective detergency requires rapid modification of surface properties to enable efficient spreading and displacement of hydrophobic soils. Although SAF shares a general structural motif with LAS, the po-

lar furan ring introduces differences that may significantly alter dynamic wetting behaviour. Therefore, quantifying the rate at which SAF solutions reduce the contact angle on representative greasy substrates, such as beef lard, provides a critical benchmark for assessing its practical utility relative to industrial standards.

In light of these gaps, this study aims to deliver an assessment of SAF to evaluate its suitability as a renewable alternative to LAS. To achieve this, the HLB of SAF will be determined experimentally using an emulsion substitution approach. Building on this characterisation, the study will profile the viscosity behaviour of SAF solutions across varied concentrations and temperatures to identify critical structural transitions. This rheological data will be corroborated by polarised light microscopy (PLM) to visually classify lyotropic liquid crystalline phases, such as hexagonal and lamellar structures, thereby allowing for the construction of a binary phase diagram. Finally, the practical applicability of the surfactant will be evaluated by quantifying the dynamic wetting performance on hydrophobic soils using sessile drop contact angle analysis. Throughout these assessments, sodium dodecyl sulphate (SDS) and sodium dodecyl benzene sulphonate (SDBS) are employed as comparative benchmarks for SAF-12 to isolate the influence of the furan based head group, given the equivalent alkyl chain lengths of these surfactants.

2 Methods

2.1 Materials and Equipment

The SAF surfactants utilised were synthesised by the Hallett Group at Imperial College London. SAF-12 was the specific species used for all experiments. Reference surfactants, sorbitan monostearate (Span 60, 45 - 55%), sorbitan monolaurate (Span 20, $\geq 44\%$), polysorbate 20 (Tween 20, $>40\%$) and polysorbate 60 (Tween 60, 40 - 60%), were also obtained from Sigma Aldrich. Target oils, hexadecane (99%), 1-octanol (99%) and 1-dodecanol (98%) were obtained from Sigma Aldrich. Rapeseed oil was obtained from a local supermarket.

For surface wetting experiments, cotton soiled with beef lard was utilised as the hydrophobic sub-

strate. All solutions were prepared using deionised water.

Viscosity measurements were performed using the ViscoQC 300 (Anton Parr). To ensure torque values remained within the instrument's operational limits, specific spindle geometries were selected based on sample viscosity: the SC4-18 spindle was utilised for concentrations of 20 - 30% (w/w), while the SC4-27 spindle was employed for the higher viscosity range of 30 - 70% (w/w).

Surface wetting was analysed using a Ramé-hart Model 590 Goniometer. Microscopic analysis was performed on a Leica DM2500 microscope equipped with polarised light filters.

2.2 HLB Determination

The hydrophilic-lipophilic balance (HLB) characterisation was conducted in two stages: determining the required HLB of the target oil, followed by the determination of the experimental HLB of the SAF surfactant. All HLB and required HLB values for surfactant blends were calculated using the weighted average method shown in Equation 1:

$$HLB_{mix} = w_A \cdot HLB_A + w_B \cdot HLB_B \quad (1)$$

where w_A and w_B are the mass fractions of surfactant A and B in the blend, respectively.

2.2.1 Required HLB of Oil Phase

To identify the required HLB for the oil phase, a series of oil-in-water (O/W) emulsions were formulated. Each emulsion consisted of 20% (w/w) oil, 4% (w/w) surfactant mixture and 76% (w/w) deionised water. Surfactant blends were prepared using Span 20, Span 60, Tween 20 and Tween 60 to span a defined HLB range by adjusting their mass ratios. The mixtures were heated to 60°C and stirred on a hot plate using a magnetic stirrer.

The aqueous and oil phases were combined by agitation using a VWR Collection Vortex Mixer for 20 seconds and then left heated at approximately 55°C. Emulsion stability was assessed by continuous observation until a clear stability trend emerged. The goal was zero phase separation; if this was not achieved with the initial surfactant ratios, additional combinations were prepared at smaller inter-

vals surrounding the least separated concentration ratio. This process was repeated until a ratio yielding zero phase separation was obtained.

2.2.2 Experimental HLB of SAF

Using the same mass ratio method employed to obtain the required HLB of the oils, the SAF surfactant (unknown HLB) was blended with a surfactant of known HLB (Span 60). These blends were then used to emulsify the target oils using the standard formulation (20% oil, 4% surfactant blend, 76% water). The objective was again to achieve zero phase separation; the same iterative procedure described previously was followed until the mass ratio that produced a stable emulsion was identified. The ratio that resulted in emulsion stability was recorded, and the HLB of SAF was subsequently calculated using Equation 1, based on the required HLB of the oils determined in the previous step and the known HLB of Span 60.

2.3 Surface Wetting

Wetting properties were evaluated via the sessile drop method on beef lard soiled fabric. Solutions were prepared at concentrations of 0.1% and at three times the CMC. For reference, the corresponding concentrations for the three surfactants are presented in Table 1.

Table 1: Concentration values at three times the CMC for SAF, SDBS and SDS (Al Ghatta et al., 2022).

Surfactant	3x CMC (w/w%)
SAF	0.1563
SDBS	0.1548
SDS	0.7085

Solutions of reference surfactants (SDS and SDBS) and deionised water were also tested following the same methodology. A droplet of 4 μ L of the surfactant solution was gently deposited onto the fabric surface using a precision syringe. The contact angle formed between the droplet and the fabric was recorded over a period of 60s (for surfactants) and 100s (for water) using a camera and ADVANCE Software from KRÜSS. The reduction in contact angle over time was analysed to quantify the wetting efficiency of the SAF surfactant relative to the other

tested solutions. Each reading was repeated for a total of three sets of data that were then averaged.

2.4 Viscosity Testing

The rheological properties of the SAF surfactant in solution were evaluated to understand their flow behaviour under varying conditions. Aqueous solutions were prepared at concentrations ranging from 20% to 70% (w/w).

Viscosity measurements were obtained using the viscometer by applying a shear force sufficient to maintain the instrument between 20–80% of its torque capacity, achieved by adjusting the spindle rotation speed (rpm) at controlled temperatures of 25 - 70°C. The data was analysed to determine the relationship between surfactant concentration, temperature, and bulk viscosity.

2.5 Polarised Light Microscopy (PLM)

To consolidate the characterisation of the phase behaviour of the SAF surfactant with the harder to identify phases, aqueous samples of SAF were prepared at 2 concentrations to correspond with the expected phases; hexagonal - 55%, lamellar - 60%. The samples and slides were heated to temperatures exceeding 40°C before microscopic examination to prevent transition into the crystalline phase.

A small quantity of each sample was placed between a glass slide and a cover slip. The samples were examined under a polarised light microscope at magnifications of 10x, 20x and 40x.

3 Results and Discussion

3.1 Interfacial & Wetting Properties

3.1.1 HLB Characterisation

Suitability of Target Oils

Emulsion stability screening was performed across four oils: hexadecane, 1-octanol, 1-dodecanol, and rapeseed oil. Hexadecane and 1-octanol proved unsuitable for this experiment, as formulations using these oils exhibited rapid and complete phase separation across the entire range of formulations for Span 20, Span 60, Tween 20 and Tween 60.

This universal instability suggests that the required HLBs for hexadecane and 1-octanol lie outside any measurable hydrophilic-lipophilic spectrum

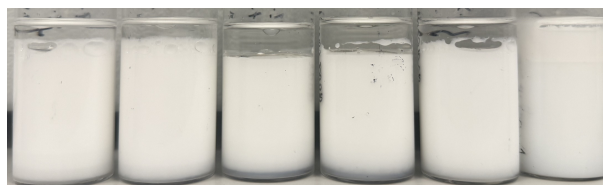


Figure 2: Visual stability assessment of a selection of 1-dodecanol emulsions formulated with Span 20/Tween 20 blends (varying linearly from 27.5% to 40% Span 20, left to right). The sample formulated with a 40% Span 20 / 60% Tween 20 ratio (rightmost) was identified as the most stable, exhibiting minimal phase separation, thereby defining the oil's Required HLB.

defined by the reference pairs. Conversely, stable oil-in-water emulsions were successfully generated using 1-dodecanol and rapeseed oil, confirming these phases as suitable for the determination of their respective required HLB values and the subsequent characterisation of the SAF surfactant.

Required HLB of Target Oils

The final stability assessment revealed distinct optimal HLB values for the compatible oils. This was achieved through bracketing experiments, with an example shown for 1-dodecanol using Span 20 and Tween 20. Figure 2 shows a partial representation of the 1-dodecanol formulations. Instability was qualitatively characterised by the number of visible phases and the size of these phases.

For the 1-dodecanol system, the optimal stability was found at a composition of 40% Span 20 / 60% Tween 20. Using Equation 1, this corresponds to:

$$\text{Required HLB} = (0.4 \times 8.6) + (0.6 \times 16.7) \approx 13.5$$

The required HLB values for both successful oil systems are presented in Table 2.

Table 2: Summary of required HLB values determined for SAF, SDBS, and SDS. The table presents the optimal surfactant blend compositions required to stabilise emulsions of 1-dodecanol and rapeseed oil, alongside the resulting required HLB values for each oil.

Oil Phase	Reference Surfactant	Optimal Blend	Required HLB
1-Dodecanol	Span 20 Tween 20	40% 60%	13.5
Rapeseed Oil	Span 60 Tween 60	87.5% 12.5%	6.0

HLB of Surfactants and Comparative Analysis

Using the substitution method, experimental HLB values were derived for SAF, as well as for the anionic benchmarks SDS and SDBS. Each surfactant was tested against both rapeseed oil and 1-dodecanol to cross validate the results. Figure 3 shows an example sequence of vials used to determine the HLB of SDS using Span 60 and rapeseed oil.



Figure 3: Visual stability assessment of a selection of rapeseed oil emulsions formulated with SDS/Span 60 blends (varying linearly from 2% to 8% SDS, left to right). The sample formulated with a 3% SDS / 97% Span 60 ratio was identified as the most stable, exhibiting minimal phase separation, thereby defining the HLB of SDS.

The results, summarised in Table 3, indicate that the determined HLB for SAF is consistent across both oil systems, falling within the range of 39.50 - 41.84. The non-linear uncertainty associated with these values was determined as half the size of the final incremental HLB step used during the bracketing procedure, reflecting the resolution limit of the visual analysis of stability.

Table 3: Summary of experimental HLB values determined for SAF, SDBS, and SDS. The table presents the optimal surfactant blend compositions required to stabilise emulsions of 1-dodecanol and rapeseed oil, alongside the resulting calculated HLB values for each surfactant.

Surfactant	Oil Phase	Optimal Blend (% Span 60)	HLB
SAF	1-Dodecanol	75	39.50
	Rapeseed Oil	96.5	41.84
SDBS	1-Dodecanol	20	15.58
	Rapeseed Oil	90	17.70
SDS	1-Dodecanol	82.5	54.41
	Rapeseed Oil	97	48.03

The theoretical HLB values for SDS and SDBS were calculated using Davies' group contribution method (Davies, 1957). This approach quantifies the HLB of a molecule by summing the distinct contributions of its constituent hydrophilic and lipophilic functional

groups, according to Equation 2:

$$HLB = 7 + \sum_i (n_i \cdot H_i) - \sum_j (m_j \cdot L_j) \quad (2)$$

where H_i represents the specific group number assigned to the i -th hydrophilic functional group, and L_j represents the group number of the j -th lipophilic group. The variables n_i and m_j denote the stoichiometric quantity of each respective group present within the molecule.

The data in Figure 4 confirms the precision of established experimental methods. The experimental HLB values obtained from literature for SDS (39.70) and SDBS (10.70) show minimal deviations of only 0.75% and 0.94%, respectively, from the theoretical benchmarks (Shi et al., 2019).

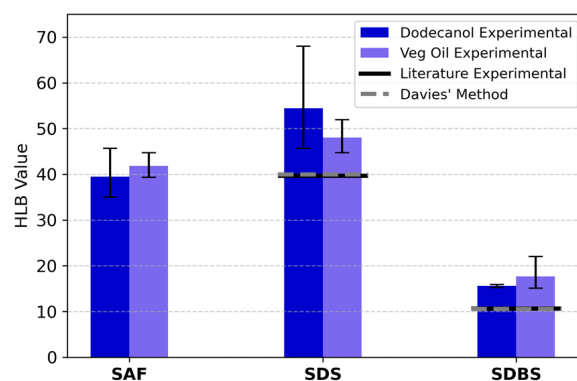


Figure 4: Comparative analysis of experimental and theoretical HLB values. The bar chart contrasts the experimental HLB values derived for SAF, SDS, and SDBS against theoretical values calculated via Davies' group contribution method and experimental values from literature. Error bars represent the uncertainty range derived from the resolution limits of the bracketing increments used during the emulsion screening process.

In contrast, the average HLB values obtained in this experiment for SDS (51.22) and SDBS (16.64) exhibit extreme deviations of 28.05% and 56.98%. The clear comparison confirms that the HLB discrepancies are primarily attributable to the methodology employed in this laboratory experiment, which failed to reproduce the accuracy achieved by established literature experimental methods.

Analysis of Experimental Error

The failure of the experimental HLB values for the known surfactants (SDS and SDBS) to align with reliable theoretical benchmarks likely stems from compounding limitations in the emulsification process.

The experiment was constrained by limitations on

the amount of surfactant available to be used. This was particularly true for SAF, which required conservative usage because of its sophisticated production process in-house. This resulted in small sample masses (20g total of water, oil and surfactant blend) and the adoption of an inconsistent, adaptive method. Initial surfactant blends were screened using large compositional increments (20%), and further narrowing to finer steps was immediately halted as soon as a single, visually stable composition was identified. This means that the determined optimal composition is fundamentally an approximation based on the closest tested blend, directly contributing to the inconsistent uncertainty ranges.

Additionally, the criteria for identifying a stable emulsion relied entirely on a qualitative visual observation of stability, despite the creation of a large number of distinct vials required for the wide sweep of formulations (four oils, four surfactants, numerous surfactant ratios). Objective, quantitative methods (such as mean droplet size measurements) became infeasible, since it was necessary to assess many formulations simultaneously to make comparisons at one particular time. Without any quantitative data, the final HLB value remains a subjective approximation, dependent on one's ability to visually discern marginal differences in phase separation.

The HLB discrepancies were potentially also exacerbated by temperature changes during sample preparation. Although stability was intended to be assessed at room temperature, the blend solutions were initially mixed and heated on a hotplate at 60°C to ensure complete dissolution.

3.1.2 Surface Wetting Performance

Figure 5 shows that all surfactant solutions reduced the contact angle relative to water alone. This is expected because the beef lard coated cotton is highly hydrophobic and normally resistant to spreading, meaning the contact angle of pure water remains high and decreases only very slowly. In contrast, the surfactant solutions contain molecules with hydrophobic tails and hydrophilic heads; these adsorb rapidly at the lard–water interface. As the hydrophobic tails interact favourably with the lard layer they reduce the interfacial tension, this allows the solu-

tion to become more compatible with the surface, allowing it to penetrate into the fabric. This results in a pronounced decrease in contact angle and enhanced wetting compared with water. The structure of the surfactant plays a crucial role in the wetting ability of the solution; surfactants with stronger hydrophobic tails will have an increased affinity to the lipid and will integrate into the layer more effectively which is a demonstration of improved wetting capability.

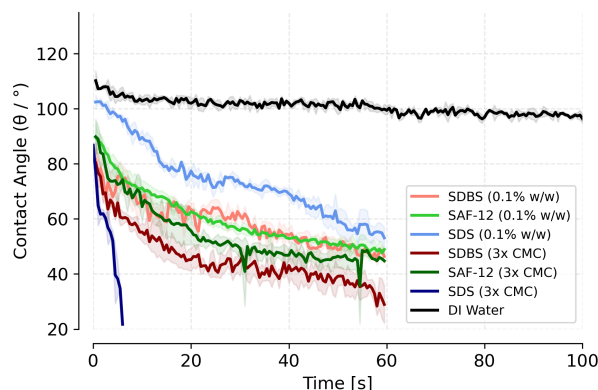


Figure 5: Dynamic wetting kinetics of surfactant solutions on beef lard soiled cotton fabric. The graph tracks the reduction in contact angle for surfactant solutions (SAF-12, SDS, and LAS at 0.1% w/w and 3x CMC) and deionised water. Data is presented as the mean of three repeats; shaded error bands correspond to half of the range of contact angle values at each time point.

At 0.1% (w/w) concentration, the SAF-12 solution exhibits a contact angle profile very similar to that of SDBS, both of which are consistently lower than the contact angle observed for SDS. This indicates that, at this absolute concentration, SAF-12 and SDBS have a higher wetting capability compared with SDS. Faster absorption suggests the surfactant solution has a higher capability to reduce the interfacial tension at the boundary of the lipid coating and allow the water to move through into the cotton fibres below. One of the key factors influencing this result is that SDS is the only surfactant that remains below its CMC at 0.1% (w/w). This indicates that, at this low concentration, SAF-12 is much more effective at wetting, per gram, than SDS.

At three times their respective CMC values, SAF-12 and SDBS show a much slower rate of reduction in contact angle compared to SDS, indicating that they have a lower penetrating rate at this concentration. However, due to the three surfactants having different CMCs, a direct comparison of their be-

haviour at these chosen concentrations must take into account that the CMC of SDS is much higher. This means that the absolute concentration of SDS when at three times the CMC is considerably higher than both SDBS and SAF, this contributes to the faster drop in contact angle and the rapid absorption of the droplet (within 10s). SAF's much lower CMC results in fewer free monomers at the same relative concentration which will slow the wetting kinetics as fewer monomers are adsorbed onto the surface in the initial stages of the drop test (Jarrahian et al., 2010). Despite this, SAF-12 still exhibits similar wetting performance to SDBS at this higher concentration. Overall, this suggests that SAF demonstrates comparable wetting ability to SDBS at both low and high concentrations, making it a viable alternative in industry.

Analysis of Experimental Error

The soiled fabric surface was not completely flat, which made it difficult for the software and camera to consistently identify the true baseline of the droplet (Figure 6 shows the blurred appearance of the fabric surface). As a result, the system occasionally misinterpreted the surface position, leading to inaccurate contact angle readings. This explains the fluctuations observed in the contact angle measurements (Figure 5). However, since five sets of readings were taken and averaged for each data point and a consistent overall trend was still apparent, the comparative results between the surfactants can be considered reliable.

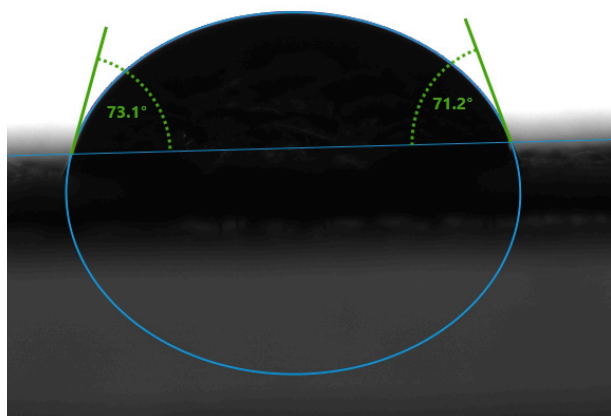


Figure 6: Representative image of a sessile drop of 3x CMC SAF solution at 20°C on the beef lard soiled fabric substrate.

3.1.3 Synthesis of Experimental Results

Evaluating the obtained HLB values and wetting behaviour highlights that SAF-12 occupies a unique position within the three surfactants analysed, effectively bridging the performance characteristics of SDS and SDBS. In terms of the HLB, SAF demonstrates a high degree of hydrophilicity, quite similar to SDS. Both surfactants yielded high experimental HLB values, with SAF ranging between 39.50 and 41.84 and SDS exhibiting similarly elevated values, in contrast to the significantly lipophilic profile of SDBS. This indicates that the oxygen heteroatom within the furan moiety contributes substantially to the polarity of the head group, ensuring high water solubility. However, the dynamic wetting behaviour of SAF-12 diverges significantly from the SDS benchmark when evaluated on a mass equivalent basis. At a fixed concentration of 0.1% w/w, SAF-12 maintained effective wetting capability (comparable to SDBS), and significantly outperformed SDS, which failed to wet the substrate effectively. While SDS requires a much higher mass concentration to reach its critical micelle concentration and activate its wetting potential, SAF-12 is capable of delivering robust surface tension reduction at a fraction of that mass. This performance suggests that the furan ring provides surface interaction properties similarly to the benzene ring of SDBS, allowing for efficient adsorption at low absolute concentrations. This combination of properties poses a distinct formulation advantage. SAF provides the superior water solubility associated with high HLB surfactants while retaining the potent, low concentration detergency typical of alkylbenzene sulphonates. Consequently, SAF emerges as a highly efficient candidate for formulation, capable of delivering robust cleaning performance without high concentration requirements.

3.2 Phase Behaviour

The phase behaviour of the SAF-water system was mapped by independently analysing the viscosity behaviour and microscopic textures, before correlating these results to construct a binary phase diagram.

3.2.1 Viscosity Behaviour

The viscosity of SAF solutions was evaluated across a concentration range of 20 - 70% (w/w) and temperatures spanning 25 - 70°C. Figure 7 illustrates the relationship between viscosity, concentration, and temperature.

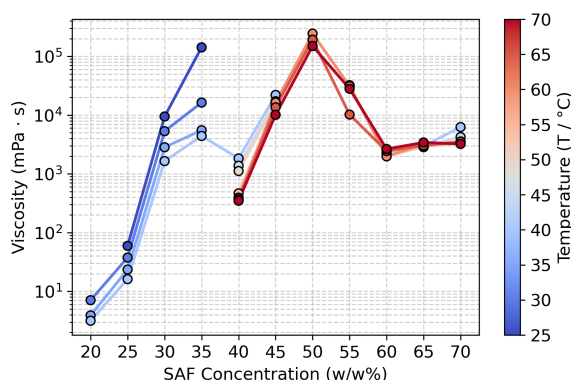


Figure 7: Rheological profile of aqueous SAF solutions across a concentration range of 20 - 70% (w/w). The plot illustrates the relationship between viscosity and concentration for isotherms spanning 25 - 70°C. Breaks in isotherms indicate transitions to solid crystalline phase where viscosity could not be measured."

The Krafft point is a temperature dependent property that defines the solubility limit of a surfactant. Below the Krafft point, the surfactant exists predominantly in a hydrated crystalline state, in which it is poorly soluble. The temperature range for viscosity measurements was restricted at different concentrations as the solutions transitioned into a solid crystalline phase at lower temperatures. This physical limitation is the reason that the isotherms in Figure 7 are not all continuous and appear to break at different points.

All tested solution concentrations were above the CMC, ensuring the presence of micelles even at the lowest concentration. A continuous increase in viscosity was observed within specific temperature-concentration bands (specifically at temperatures below 40°C for concentrations 25 - 35% and temperatures more than 55°C for concentrations 40 - 50%). This trend reflects the growth and structural evolution of the micelles; as concentration rises, spherical micelles elongate into oval and eventually worm-like structures. The increasing entanglement among these aggregates restricts flow, resulting in the observed viscosity increase.

A critical rheological transition is observed at 55%

for temperatures above 55°C, demonstrated by a sudden, sharp drop in viscosity. The onset of this region was also observed at 40% concentration at 45°C. Due to the heightened concentration, the micelles here are more densely packed and interact further, forming indefinite rod-like tubular structures arranged in a two-dimensional array.

Viscosity drops further at 60% (for temperatures above 55°C) before slowly increasing again for values greater than 60%. This behaviour marks the transition into the lamellar phase. The layered organisation of this high concentration phase results in a decrease in viscosity due to the ability of adjacent surfactant bilayers to shear and slide over one another, thereby facilitating enhanced fluidity (Hendrikse, 2022).

3.2.2 Microscopic Phase Identification

To identify the specific mesophases responsible for the rheological transitions, PLM was performed.

At the concentrations corresponding to the initial viscosity drop, the sample exhibited a distinct texture under polarised light. As shown in Figure 8, the structure displays a "mosaic-like" pattern. This optical texture is characteristic of the hexagonal phase, confirming that the micelles at this concentration pack into hexagonally arranged cylinders (Hendrikse, 2022).

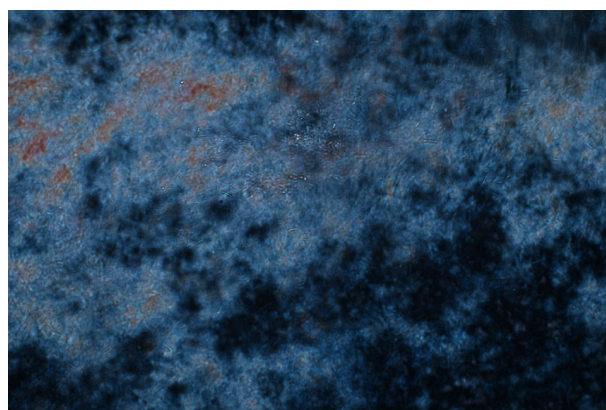


Figure 8: PLM micrograph of 50% (w/w) aqueous SAF solution at 40x.

At a concentration of 60%, the texture under the microscope shifted distinctly, confirming the transition to the lamellar phase. Samples displayed characteristic Maltese cross patterns at 10x magnification, and smokey streaks at 40x magnification, as

shown in Figure 9. These textures are diagnostic of the stacked arrangement of surfactant bilayers separated by aqueous layers of uniform thickness (Hendrikse, 2022).

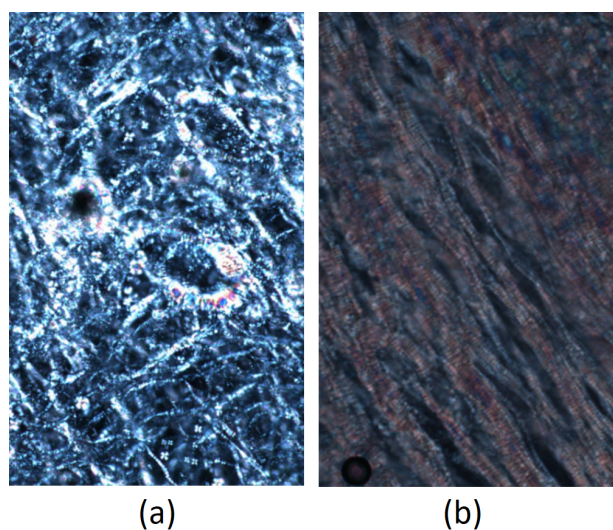


Figure 9: PLM micrographs of 60% (w/w) aqueous SAF solution at (a) 10x and (b) 40x magnification.

3.2.3 Phase Diagram Construction

By correlating the temperature and concentration dependent viscosity limits with the confirmed microscopic textures, a partial binary phase diagram for the SAF-water system was constructed (Figure 10).

The boundaries were defined by overlaying the rheological transition points with the visual confirmation of certain phases. The observed phase progression aligns with the typical sequence established for standard anionic surfactants such as SDS, evolving from a micellar solution at varied temperatures and concentrations into a hexagonal phase and finally transitioning to a lamellar structure at concentrations equal to or exceeding 60% (Dierking and Al-Zangana, 2017).

Temperature plays a critical role in the phase behaviour. At sufficiently low temperatures, the system consistently shifts into a crystalline phase. This occurs because decreasing temperature reduces molecular motion, allowing the surfactant tails to pack closely into an ordered, energetically favourable arrangement. The resulting phase is solid and brittle, and typically appears paler than the higher temperature liquid crystalline and micellar phases (Hendrikse, 2022).

To improve the resolution of the phase diagram, future work should include viscosity measurements at smaller concentration and temperature intervals to precisely map the coexistence regions between phases.

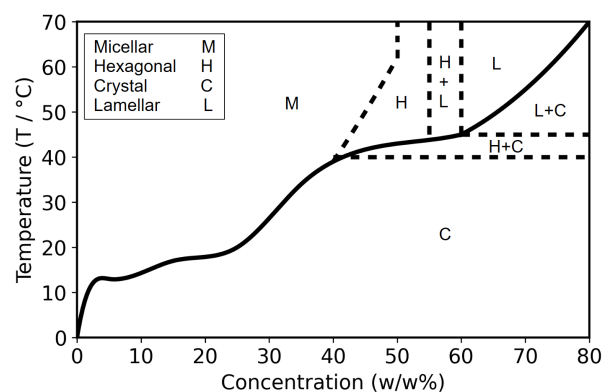


Figure 10: Proposed binary phase diagram for the SAF-water system. The diagram maps the regions of micellar, hexagonal, lamellar, and crystalline phases along with the expected transitional phases. Boundaries were constructed by correlating the rheological transitions observed in Figure 7 with the specific liquid-crystalline textures identified via PLM analysis.

4 Conclusions

This study aimed to characterise the phase properties, wetting behaviour and HLB of SAF. The key findings were that SAF had an experimentally determined HLB of 40.67, its wetting behaviour had a similar penetration rate to SDBS above and below the CMC, and its phase behaviour revealed four distinct phase regions with potential transitional phases likely to appear with finer concentration and temperature resolution.

The results show that SAF behaves as an effective anionic surfactant, with its average experimental HLB value of 40.67 indicating a high degree of hydrophilicity, making it suitable for application as an oil-in-water emulsifier or detergent. Its wetting performance, comparable to SDBS, suggests it can adequately reduce surface tension for detergent applications. The distinct phase regions observed indicate stable mesophase formation, with more possible transitional phases hinting at additional formulation flexibility.

The key limitations of this study were as follows: the HLB determination relied on subjective visual assessment of the phase separation and the small ratio increments may have reduced precision, the wetting

tests were affected by inconsistent beef lard saturation on the fabric, and the polarised light microscope could not be temperature controlled meaning unintended phase transitions may have occurred during imaging.

Further work should include finer phase mapping across concentration and temperature ranges, testing SAF with additional oils and fabrics, and obtaining more micrographs in a controlled temperature setting. Linking the surfactant's HLB and phase behaviour to real formulation performance would also help clarify its practical applications.

Acknowledgements

The authors of this report would like to acknowledge Bo Gu for his valuable support throughout the duration of this research. We would also like to extend our thanks to the Hallett Research Group for their guidance and resources throughout this project.

References

- Al Ghatta, A., Aravenas, R. C., Wu, Y., Perry, J. M., Lemus, J., and Hallett, J. P. (2022). New biobased sulfonated anionic surfactants based on the esterification of furoic acid and fatty alcohols: A green solution for the replacement of oil derivative surfactants with superior proprieties. *ACS Sustainable Chemistry & Engineering*, 10(27):8846–8855.
- Davies, J. (1957). A quantitative kinetic theory of emulsion type, i. physical chemistry of the emulsifying agent. In *Gas/Liquid and Liquid/Liquid Interface. Proceedings of the International Congress of Surface Activity*, volume 1, pages 426–438.
- Dierking, I. and Al-Zangana, S. (2017). Lyotropic liquid crystal phases from anisotropic nanomaterials. *Nanomaterials*, 7(10):305.
- Hendrikse, R. L. (2022). *The Phase Structure and Rheological Properties of Surfactant Solutions via Dissipative Particle Dynamics*. PhD thesis, University of Leeds.
- Jarrahian, K., Vafie-Sefti, M., Ayatollahi, S., Moghadam, F., and Moghadam, A. (2010). Study of wettability alteration mechanisms by surfactants.
- Kaida, H., Syed, M., Shukor, Y., and Othman, A. (2021). Biodegradation of linear alkylbenzene sulfonates (las): A mini review. *Bioremediation Science and Technology Research*, 9:1–6.
- Scheibel, J. (2004). The evolution of anionic surfactant technology to meet the requirements of the laundry detergent industry. *Journal of Surfactants and Detergents*, 7:319–328.
- Shi, Y., Yan, F., Jia, Q., and Wang, Q. (2019). Norm descriptors for predicting the hydrophile-lipophile balance (hlb) and critical micelle concentration (cmc) of anionic surfactants. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 583:123967.

The Use of Neural Networks in the Prediction of Thermal Conductivity of Organic Liquids: Key Features and the Added Value of Molecular Fingerprints

Lynn Berri and Ali Hayat

Department of Chemical Engineering, Imperial College London, U.K.

Abstract: This study investigates the contribution of molecular fingerprints relative to traditional physical descriptors by comparing an ANN model using Morgan fingerprints derived from SMILES with a numerical-only baseline. On the validation set, the full model containing the morgan fingerprints achieved a WMSE of 0.000309, significantly outperforming the numerical-only model (0.0005497), as confirmed by a Wilcoxon test. Forward feature addition identified a compact model consisting of SMILES + Temperature + Density + Dipole moment + Acentric factor + Boiling point, which achieved a validation WMSE of 0.000333. A Wilcoxon test showed no statistically significant difference between this reduced model and the full model. These results demonstrate that Morgan fingerprints supply structural information beyond coarse numerical descriptors while requiring only a small descriptor subset to match full-model performance.

1 Introduction

The search for new heat transfer liquids is increasingly driven by both environmental and performance demands [1,2]. Sustainable, non-toxic, and biodegradable fluids are required to meet tightening environmental regulations, while advanced thermal systems impose physical demands for high efficiency, material compatibility, and operational reliability [3]. A key example of this is the use of energy in data centres where by 2030, 9% of annual US electricity generation will be consumed by data centres. Currently, 40% of data centre annual energy consumption is related to cooling systems [4]. Accurate and efficient thermal conductivity prediction is crucial because it supports the design of safer and more efficient thermal systems and accelerates innovation in heat-transfer technology for diverse and demanding applications. For example, organic heat-transfer fluids are often used in industrial process heating systems operating at elevated temperatures, particularly in chemical processing [5]. Therefore, it is vital to obtain reliable thermal conductivity data.

Thermal conductivity, k , of liquids is a measure of its ability to conduct heat and is influenced by factors such as the liquid's molecular structure, temperature and pressure [6]. It is normally expressed using Fourier's Law as follows [7]

$$\dot{q} = -k\nabla T \quad (1)$$

where $\dot{q}$, k , and ∇T are the local heat flux (W m^{-2}), thermal conductivity ($\text{W m}^{-1} \text{K}^{-1}$), and the temperature gradient (K m^{-1}), respectively [8]. It is a key transport property governing heat transfer in organic liquids and appears explicitly in correlations, typically expressed via the Nusselt and Prandtl numbers, film condensation coefficients, and other dimensionless groups used in heat transfer design [6].

In dense liquids, heat transfer depends on collision frequency and energy transfer efficiency in collisions, which are strongly influenced by physical descriptors such as density (packing), dipole moment, dielectric constant and heat capacity, among others. These descriptors already carry structural information indirectly because molecular structure influences [9].

Experimental methods have been used to calculate k by using a transient hot wire, but detailed uncertainty analysis is required to reach $\pm 0.96\%$ at 90°C [10]. Although experimental methods can give very accurate values, reaching as small as $\pm 0.2\%$ [11], this method of obtaining k is a long and laborious process and requires careful calibration as errors can arise from wire calibration and convection onset. Data is only available for a limited set of liquids and temperature so improvements in machine learning models can be used as a replacement to accurately find k for new synthetic heat transfer fluids. The data used in the machine learning model was taken from DIPPR 801 and evaluated by an expert for each liquid, giving the model the best possible and non-conflicting data.

Correlations such as the Sastri-Rao correlation, a relationship used to predict the k of saturated organic liquids, depend heavily on the experimental data used to fit them and are limited to a specific temperature range for specific molecular groups [12].

Hence the use of Artificial Neural Networks (ANNs) in this study, a machine learning model, as they can outperform traditional semi-empirical correlations, group-contribution and molecular simulation methods in both accuracy and computational efficiency. Machine learning models provide a flexible framework to learn structure–property relationships for transport properties and have already shown promising performance for viscosity and thermal conductivity of fluids and mixtures [13]. Recent models for liquid

thermal conductivity have shown lower prediction errors and better generalization across temperatures and functional groups than correlations [14]. Literature shows machine learning models have been used to predict k values of concrete, where mean absolute errors of 0.07, 0.14 and 0.10 W m⁻¹ K⁻¹ were achieved [15]. This strong performance indicates that machine learning models can effectively capture the complex, non-linear relationships between concrete-mix parameters and thermal conductivity that traditional empirical models struggle to represent. Although concrete and organic liquids are materially very different, the underlying challenge is similar where, in both cases, thermal conductivity arises from multi-factor interactions that are not easily described by classical correlations.

The molecular structure of a liquid can be encoded as SMILES (Simplified Molecular Input Line Entry System) strings transformed into fingerprints or graph-based representations. One of the central objectives of this study is therefore: in the presence of multiple coarse physical descriptors, which already encode much of the variation in thermal conductivity and indirectly reflect structural effects, to what extent does adding explicit molecular-structure information via SMILES provide additional predictive power? In other words, how much does explicitly giving molecular structural information to the model provide additional predictive power beyond the indirect structural effect in relevant physical descriptors?

In addition, the rich feature set we explore contains descriptors that are highly correlated and partially redundant in terms of physical information they carry. A key objective is whether the ANN leverages the physical information in all these features, or whether some features are used primarily as structural surrogates while their underlying physical contributions are effectively captured by other, more explicit thermophysical descriptors. The goal is to identify which features serve mainly as structural differentiators and which continue to encode meaningful physical information beyond molecular structure alone.

The study focuses on a diverse subset of organic liquids, including simple families such as alkanes, alkenes, ketones, aromatics, as well as more complex molecules with multiple functional groups. This chemical diversity ensures that the neural networks predictions can be generalized across a wide range of organic liquids, allowing for the evaluation of both the robustness and transferability of the models.

2 Methodology

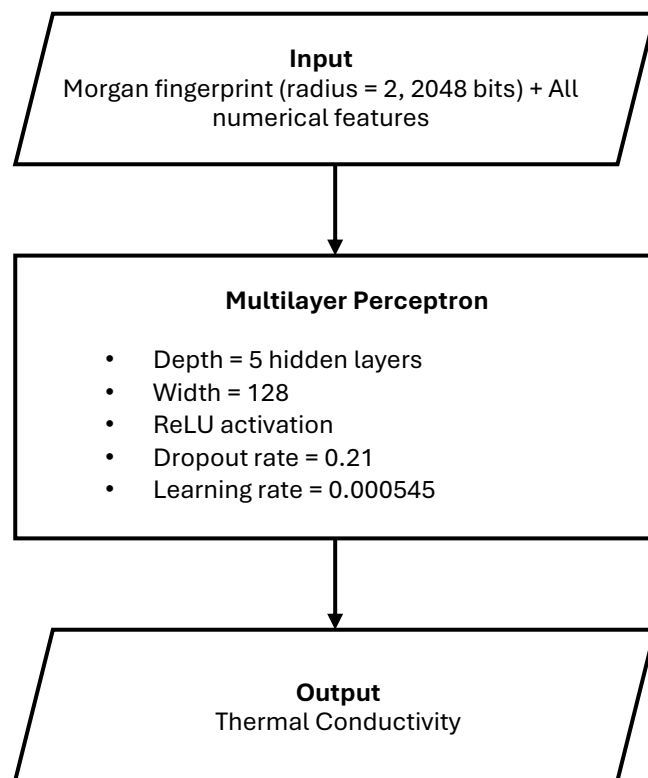


Figure 1 Baseline ANN model used and its hyperparameters

SMILES is a text-based representation of a chemical molecule that encodes its structure using a linear ASCII string [16]. To capture detailed structural information from these representations, we generated molecular fingerprints—numerical binary vectors where each bit indicates the presence or absence of a specific substructure, such as functional groups, bond types, or branching patterns [17]. In this study, we used Morgan fingerprints (also called circular fingerprints), which are constructed by iteratively exploring the local environment around each atom up to a specified radius [18]. A radius of 2 was selected to ensure that the fingerprints capture both the immediate atomic environment and second-neighbour interactions, providing sufficient specificity to distinguish between molecules with subtle differences in connectivity or branching without generating excessively large feature vectors.

The thermal conductivity prediction model is an ANN implemented using TensorFlow/Keras. The network consists of an input layer that accepts the selected features, multiple fully connected hidden layers with ReLU activation, and a single output neuron for regression. Dropout regularization was applied after hidden layers to reduce overfitting, while the Adam optimizer was used for training. The network was trained using a weighted mean squared error (WMSE) to account for varying uncertainties in the experimental data. Early stopping and custom

callbacks were included to ensure stable training and prevent divergence.

To ensure a fair and controlled comparison between feature sets, a single baseline hyperparameter configuration was used for all neural network models. One might argue that a fixed configuration could, by chance, be better suited to one feature set than another. While this is a valid concern, performing a separate hyperparameter optimisation for each feature set would introduce far more serious methodological problem: it would confound the effect of the model with the effect of the features themselves.

Hyperparameter tuning adapts the neural network architecture and learning dynamics to the specific feature set being used. If each feature set were allowed to tune its optimal hyperparameters, then differences in predictive accuracy would reflect a combination of (i) the intrinsic information content of the features and (ii) how effectively the model was tailored to them. This would make it difficult to determine whether a performance improvement arises from additional information present in the feature set or merely from more favourable hyperparameters. Moreover, hyperparameter search does not guarantee finding the global optimum, which further increases variability and injects noise in the comparison.

Using a fixed baseline hyperparameter set eliminates this confounding effect: every model is trained with the same capacity, same architecture, and same learning dynamics ensuring that any improvement in predictive performance can be attributed directly to differences in the information content of the feature sets. This approach isolates the contribution of the features themselves and provides a clean, interpretable comparison of their relative predictive value.

The concern that a single baseline hyperparameter configuration could unfairly favour one feature set over another is especially relevant in our study, where the input dimensionality and structure vary dramatically across models. This is particularly true for models that incorporate SMILES-derived fingerprints: integrating SMILES increases the feature dimensionality by 2048 bits, whereas numerical-only models operate on fewer than 20 features. Under such conditions, it becomes difficult to identify one universal hyperparameter configuration that is equally well-suited to both low-dimensional numerical descriptors and high-dimensional molecular fingerprints.

To mitigate this issue while still isolating the true effect of the feature sets, we adopted a dual phase hyperparameter strategy.

Table 1 The 19 numerical features used in the first model

The 19 numerical features
Molecular Weight
Melting Point
Boiling Point
Dipole moment

van der Waals volume
van der Waals area
Temperature
Triple Point Temperature
Triple Point Pressure
Critical Point Temperature
Critical Point Compressibility
Acentric Factor
Density
Heat capacity
Viscosity

An Optuna hyperparameter search with N = 50 trials was performed separately for each of the following feature sets:

1. SMILES + 19 numerical descriptors
2. 19 numerical descriptors
3. SMILES + (Temperature, Density)
4. Reduced numerical subset containing physically meaningful descriptors (van der Waals area, van der Waals volume, radius of gyration, heat capacity, dielectric constant, temperature, and density)

These four sets span a wide range of dimensionalities and include both purely numerical and SMILES-integrated models. From each feature-set specific search, we extracted the top five performing hyperparameter configurations based on validation performance. This produced a pool of 20 candidate hyperparameters across all four models.

Each model was then re-trained using all candidate hyperparameter configurations, and for every configuration we computed the weighted mean squared error across all four feature sets. The purpose of this step is to identify the hyperparameters that not only perform well for one specific feature set but generalize robustly across feature sets of radically different sizes and properties. The final hyperparameter configuration selected for all comparative model training was the one that minimized the average weighted mean squared error (WMSE) across all feature sets.

This ensures that:

- No single feature set disproportionately influences the hyperparameter choice.
- The chosen configuration is not biased toward either SMILES or numerical-only models.
- The performance differences observed later in the study are attributable to the feature sets themselves, not the choice of architecture.

2.1 Full model vs Numerical model

We began by comparing SMILES combined with 19 numerical features against the 19 numerical features alone. Each feature set was trained using the baseline

hyperparameters across ten seeds, and the Wilcoxon test was applied afterwards to assess statistical significance between the validation WMSEs. All 19 numerical features were included in this step to test whether, when all physical descriptors that encode variants of structural information are present, molecular fingerprints can still contribute to thermal conductivity beyond its indirect influence on these features.

The same methodology was adapted for all other model comparisons trained on the same baseline hyperparameters.

Table 2 Models trained and compared

Model Comparisons
Full model vs Numerical only model
Full model vs Filtered model
Filtered model vs Reduced filtered model
SMILES + T + ρ vs Full model
SMILES + T + ρ vs SMILES + T + $\rho^{2/3}$
SMILES + T + ρ vs SMILES + T + ρ /MW
SMILES + T + ρ + Dipole moment + Acentric factor + Boiling point

2.2 SHAP analysis of Full model vs Numerical model

We performed a SHAP (SHapley Additive exPlanations) analysis to quantify the contribution of each feature to the prediction of thermal conductivity. SHAP is a game-theory based method that assigns an importance value to each feature by comparing the model’s prediction with and without that feature, effectively measuring how much each input contributes to the output [19]. In our study, we used the model-agnostic Kernel SHAP approach, which perturbs input instances, makes predictions on these perturbed inputs, and fits a weighted local linear model to compute SHAP values [19]. We selected a sample size of 100 for both the background and validation subsets to balance computational efficiency with adequate representation. The SHAP analysis was performed for both the full model (SMILES + numerical descriptors) and the numerical-only model. By comparing the SHAP values between these two models, we aimed to identify which features lose importance when SMILES is included, thereby revealing which features were playing a structural role and could be partially or fully replaced by the structural information encoded in SMILES.

2.3 Filtered vs Reduced Filtered Model

The filtered model was created by removing Van der Waals volume, TPT, TPP and viscosity from the full model because these features lost five times their SHAP importance in the presence of SMILES.

The reduced filtered model was created by removing four descriptors from the filtered model: radius of gyration, CPC, molecular weight, and CPV.

Radius of gyration was removed not because of low SHAP importance, but because its geometric information can almost be entirely encompassed by the more important Van der Waals area; this redundancy is confirmed by their correlation of 0.96. Eliminating such correlated features allow us to simplify the dataset while preserving relevant geometric information relevant to thermal conductivity.

Molecular weight was removed for a different reason. It showed low SHAP importance in both the full model and the numerical-only model. This is likely because molecular weight does not serve as a structural proxy in these models. Many organic molecules can have the same molecular weight while differing dramatically in branching, rigidity, polarity, or functional groups.

Once SMILES is provided, the structural information that molecular weight vaguely approximates becomes explicitly available to the model. Therefore, removing molecular weight eliminates a potential weak, largely non-informative descriptor whose limited structural role can be better captured by SMILES.

However, this descriptor may still carry non-structural information relevant to thermal conductivity. For example, in Enskog theory, the number-density term ($\rho N_A/MW$) incorporates molecular mass and influences collisional transport, contributing to macroscopic transport properties [20,21]. This physically meaningful contribution is assessed later in our analysis to ensure that removing molecular weight does not inadvertently discard essential non-structural information.

2.4 Finding a Small Subset that replicates the behaviour of the full model

It is important to recognize that neural networks learn non-linear and non-additive relationships between features. This means that the individual contribution of a descriptor does not simply add up, and its usefulness can change depending on which features are present. As a result, there is no straightforward way to identify a single more optimal subset by simple ranking.

SHAP values can guide feature selection, but only to a limited extent. Adding features in descending SHAP importance may seem logical, yet a better-performing subset, because SHAP values themselves depend on the context of the full feature set. When the feature set is reduced, the same descriptor may gain or lose predictive power depending on the combination of variables which remain.

Similarly, backward elimination can be misleading if interpreted naively. Removing an important feature may initially reduce performance, but this decrease can be temporary: once certain correlated or redundant features are removed, the network may learn to leverage the remaining descriptors more effectively. This delayed recovery of performance is a

direct consequence of the nonlinear feature interactions learned by the ANN.

We therefore adopted a dual approach, performing both forward and backward selection, guided by physical intuition and SHAP-derived insights. This strategy allows us to better explore the feature space and identify reduced subsets that preserve predictive accuracy.

2.5 SMILES + T + $\rho^{2/3}$ and SMILES + T + ρ/MW transformations

To further investigate the role of molecular packing and size in thermal conductivity, we introduced two physically informed transformations of density: $\rho^{2/3}$ and ρ/MW [9, 21]

- $\rho^{2/3}$: This transformation is related to the typical scaling of surface area with volume in three dimensional objects. In liquids, it can be interpreted as a proxy for the effective collisional cross-section or the average distance between molecules. This helps the ANN capture how packing density affects collision frequency and energy transfer efficiency, which are critical for thermal conductivity.
- ρ/MW (number density): Although ρ/MW is not the exact number density (because true number density requires multiplication by Avogadro's number), it is directly proportional to it and therefore encodes the same physically relevant trend: the relative number of molecules per unit volume, which strongly influences collision frequency and thermal transport in dense liquids.

2.6 SMILES + T + ρ + Dipole moment + Acentric factor + Boiling point

For this model we stopped at the top five SHAP-ranked features (from the full model) – SMILES, combined temperature, density, dipole moment, and boiling point because together they capture the dominant physical factors that govern thermal conductivity. Thermal conductivity is primarily determined by (i) molecular structure, which sets collision geometry and energy-transfer pathways, (ii) thermodynamic state, which dictates molecular packing and collision frequency, and (iii) the strength and organization of intermolecular forces. These five descriptors align directly with these determinants: SMILES provides explicit structural information, including branching, ring systems, and functional groups; temperature and density define the thermodynamic state and packing conditions; dipole moment encodes polarity and directional interactions; and boiling point summarizes cohesive forces. Acentric factor reflects shape-induced deviations in cohesive behaviour, which neither previously mentioned descriptors capture. SHAP analysis on the full model suggested that adding numerical features beyond these five could produce diminishing returns, with descriptors such as heat capacity or critical properties providing little to or no new independent information (although this could change depending on the combination of feature sets).

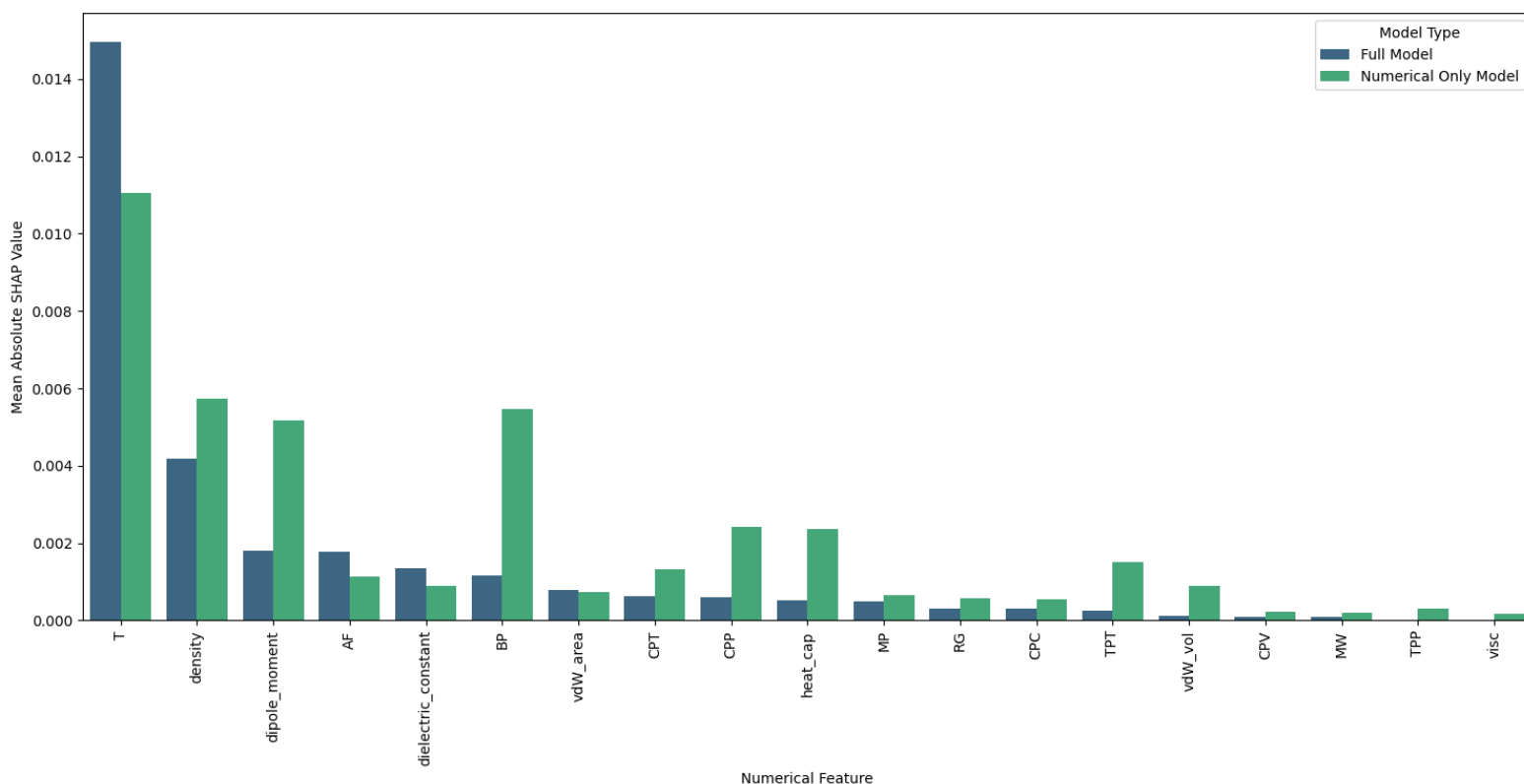


Figure 2 SHAP Analysis showing the importance of each feature in the full model and the numerical model

3 Results and Discussion

3.1 Full Model vs Numerical Model

The WMSE of the full model (SMILES + 19 numerical descriptors) was 0.000309, whereas the numerical only-model had a WMSE of 0.000547, an increase of 45%. The Wilcoxon test returned a p-value of 0.0020, indicating that this improvement is statistically significant.

This result shows that, although coarse descriptors such as density capture structural effects related to packing and dipole moment captures structural aspects of polarity and short-range interactions, adding Morgan fingerprints still provides significant predictive information. In other words, the network is using explicit local molecular connectivity patterns beyond the indirect information already present in the numerical descriptors.

Molecular structure can affect liquid thermal conductivity through several mechanisms that Morgan fingerprints help the ANN access:

1. Local topology and branching patterns. Morgan fingerprints encode the local arrangement of atoms and bonds, capturing branching versus linear chains and ring connectivity within a given radius. These local motifs influence how molecules pack and the distribution of collision geometries, even when bulk density is similar.
2. Functional-group context and relative positioning. The fingerprints encode which functional groups are present near each atom, including adjacent or conjugated groups. This information correlates with local polarity patterns and hydrogen-bonding propensity, affecting energy transfer during molecular collisions.
3. Local rigidity and degree of unsaturation. While fingerprints do not provide vibrational frequencies or 3D bond stiffness, they encode the presence of double bonds, aromatic systems, and fused rings within the local environment. These features influence local flexibility and vibrational mode distribution, indirectly affecting energy transfer.
4. Steric motifs and accessible contact geometries. Morgan fingerprints capture whether bulky or adjacent substituents are present around each atom. These topological signals correlate with steric hindrance at collision sites, influencing how efficiently momentum and energy are transferred.

By encoding these local structural features, Morgan fingerprints enable the ANN to learn the direct impact of molecular topology and local steric environments on thermal conductivity, beyond the indirect information already present in numerical descriptors.

3.2 SHAP Analysis of Full model vs Numerical Model

The SHAP analysis, seen in Figure 2, revealed that a few numerical features lost approximately fivefold in importance when SMILES was included, specifically TPT (Triple Point Temperature), viscosity, vdW volume and TPP (Triple Point Pressure). This drop in importance indicates that, in the numerical-only model, these features were primarily used by the ANN as structural identifiers to differentiate molecules rather than for their intrinsic physical information. Importantly, this does not mean the physical information they encode has been lost; rather it is now captured in other correlated descriptors. For example, TPT is perfectly correlated with melting point, which remains important because it contributes physically to thermal conductivity by reflecting the energy needed for molecular mobility and the strength of the intermolecular interactions.

Van der Waals volume, while losing importance, reflects 3D geometric constraints critical for thermal conductivity. These variants which affect the effective collision cross-section, detailed conformational extent and free-volume geometry – remain captured by other inputs (e.g., radius of gyration and vdW area) that retained importance.

Features such as heat capacity, boiling point and dipole moment experienced approximately a twofold decrease in SHAP importance. This suggests that while SMILES captures part of the structural information these features encode (e.g., branching, functional groups, torsional modes for heat capacity), their non-structural thermodynamic contributions remain relevant. For example, heat capacity reflects vibrational energy storage that cannot be inferred from other physical descriptors, so it still contributes uniquely to thermal conductivity.

Interestingly, viscosity maintained a low SHAP value in both the numerical-only and full models. Viscosity affects the momentum transfer between molecules, while thermal conductivity relates to energy transfer. In dense liquids, these properties are often correlated because both depend on molecular size, shape, and interaction strength. Theories such as the Enskog theory and kinetic theory of dense fluids explicitly link viscosity and thermal conductivity into the efficiency of energy transfer, but this contribution while theoretically supported is not shown for our data set and neural network.

3.3 Filtered Model vs Full Model

After identifying which structural patterns are explicitly encoded by the SMILES representation, we removed the corresponding physical descriptors to test whether their predictive contribution was indeed being captured by SMILES alone. We then compared

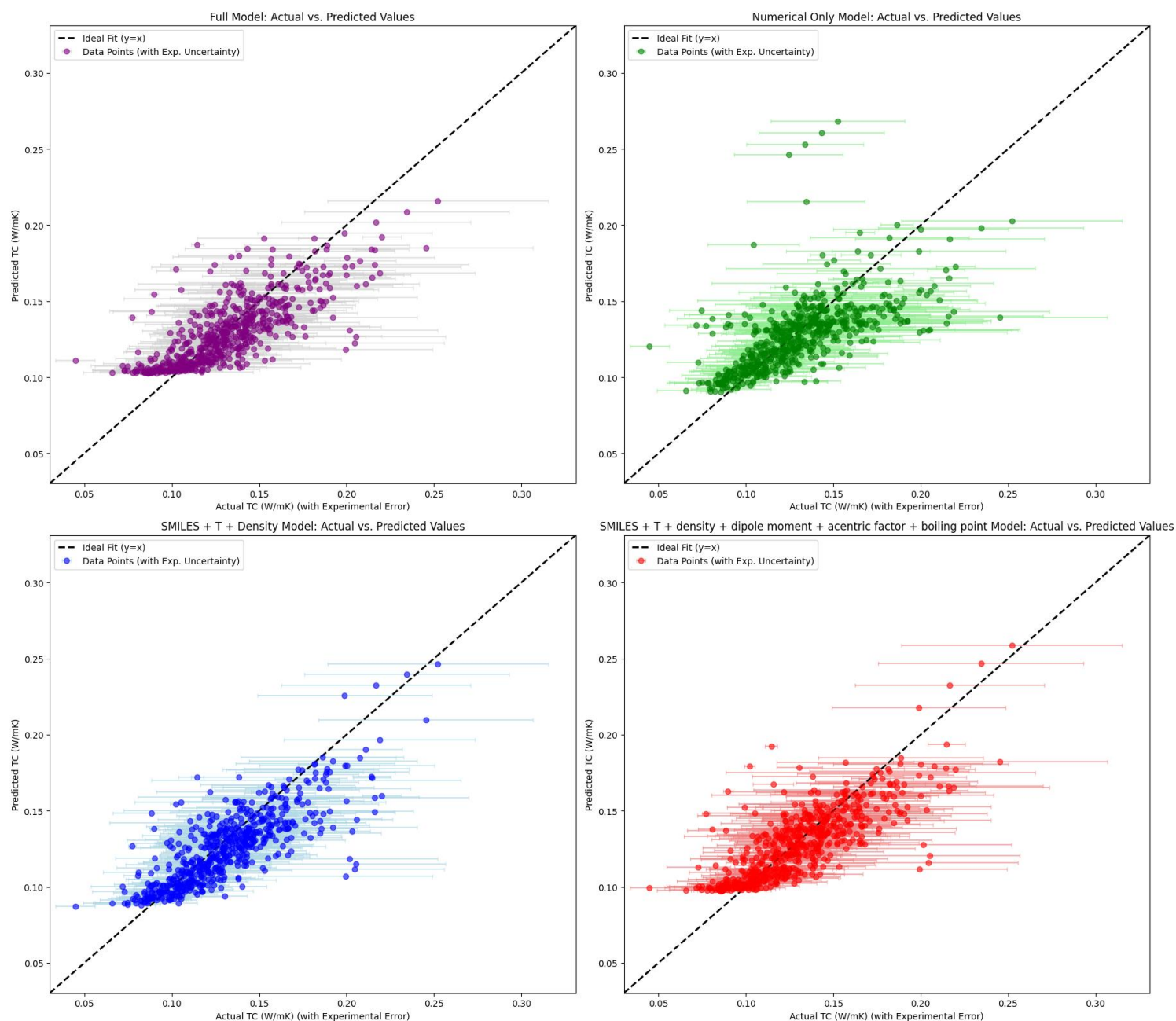


Figure 3.3 Parity plots for the 4 different models tested

the baseline and filtered model (which involved removing vdW volume, TPT, TPP, and viscosity from the full model). We then compared the baseline and filtered models using the Wilcoxon signed-rank test. The filtered model achieved a mean validation error of 0.000342, compared to 0.000309 for the baseline model. The Wilcoxon test yielded a p-value of 0.3750 indicating no statistically significant difference between the two models. This confirms that removing the structurally redundant descriptors does not degrade performance and therefore supports the validity of our SHAP-based feature interpretation.

3.4 Filtered vs Reduced Filtered Model

The reduced filtered model achieved a mean WMSE of 0.000368, compared to 0.000342 for the filtered model – an increase of approximately 7.6%. The Wilcoxon test

confirmed that this difference is statistically significant. This indicates that the removed features, although they exhibited low SHAP importance and were highly correlated with other descriptors, still contributed measurable predictive power when embedded within the larger feature set.

To investigate whether accuracy drop was driven by a specific descriptor, we reintroduced each removed feature individually into the reduced filtered model and retrained it. If a single feature were responsible for the lost accuracy, its reintroduction should have restored performance. However, no individual feature, led to statistically significant improvement, implying the lost accuracy arises from the synergistic interaction of the removed features rather than any single descriptor.

However, this decrease is not definitive evidence that these features are required; because neural

networks learn nonlinear, non-additive feature relationships, the effect of removing features can be temporary and dependent on the specific combination of remaining descriptors.

3.5 Comparison of the Transformed Features Models

SMILES + T + ρ vs SMILES + T + $\rho^{2/3}$ vs SMILES + T + ρ/MW

The SMILES + T + Density Model achieved a mean validation WMSE of 0.000427, representing a substantial decrease in accuracy compared to 0.000309 of the full model. This strongly indicates that thermal conductivity cannot be reliably inferred from packing (capture through density), thermodynamic state (capture through temperature), and molecular-connectivity information alone. A lot of group contribution and semi-empirical correlations for liquid thermal conductivity are built around these descriptors mentioned: bulk thermodynamic state (T), sometimes density or reduced density, and coarse structure via groups or corresponding-states parameters. These models work reasonably well on average but typically show mean deviations of several percent and degrade when extrapolated in chemical space. They implicitly assume that “packing + a few coarse structural indicators” capture most of the variance in thermal conductivity. Thus, the result highlights the additional information our network can exploit.

Both density transformations, $\rho^{2/3}$ and ρ/MW , did not produce statistically significant improvement in predictive accuracy when compared directly to the baseline model using SMILES + temperature + density. This outcome suggests that these transformed descriptors do not introduce new, independent physical information beyond what is already inferred from the raw density. The ANN can infer the relevant packing-and collision-related scaling relationships from density itself, making additional manipulations such as characteristic length scaling ($\rho^{2/3}$) or approximate number-density-scaling (ρ/MW) redundant. As a result, either transformation alters the model’s predictive capability in a meaningful way for this dataset. (exact mean validation WMSE values to be inputted)

3.6 Best Performing Model

The subset SMILES + temperature + density + dipole moment + acentric factor + boiling point achieved a mean WMSE of 0.000333, compared to 0.000309 for the full model. The Wilcoxon test revealed that this difference is not statistically significant. This indicates that these five numerical descriptors combined with SMILES capture the majority of physically meaningful variance in thermal conductivity for the studied organic liquids, effectively replacing the need for the remaining 13 features.

While this subset effectively captures the dominant structural and physical determinants of thermal conductivity. Other combinations of descriptors could achieve comparable accuracy, highlighting that multiple minimal subsets may encode the key physical information. In this study, the choice was guided by both SHAP importance and physical interpretability.

3.7 Comparison of the Best Performing Model against the Full Model on the test set

We then took the baseline full model and compared it against the reduced feature subset consisting of SMILES + temperature + density + dipole moment + acentric factor + boiling point, training both models using the same baseline hyperparameters as before on the test set. The full model achieved a mean test WMSE of 0.000397, whereas the reduced subset achieved a mean test WMSE of 0.000380. A Wilcoxon signed-rank test on the test predictions confirmed that this difference is not statistically significant ($p = 0.857$). This demonstrates that the reduced feature set offers replicable and statistically equivalent performance relative to the full model. Thus, the combination of SMILES, temperature, density, dipole moment, acentric factor, and boiling point is sufficient and generalizable, capturing the essential physical variation required for accurate thermal conductivity prediction without the need for the remaining 13 descriptors.

4 Conclusion and Outlook

In conclusion, this study explored the determinants of thermal conductivity in organic liquids using artificial neural networks and the added value of SMILES-derived molecular fingerprints. We found that while coarse numerical descriptors such as density, dipole moment, and dielectric constant capture indirect structural and physical effects, the inclusion of Morgan fingerprints significantly improves predictive accuracy by 45%, indicating that finer-grained local connectivity, functional-group context, branching, and steric motifs play a critical role in determining thermal conductivity. Feature selection analysis revealed that a reduced subset containing SMILES, temperature, density, acentric factor, dipole moment, and boiling point achieves performance comparable to the full model, suggesting that these descriptors effectively capture the most important physical and structural variants. Transformations of temperature and density did not yield additional predictive power. Looking forward, physics-constrained neural networks and graph-based models like GNNs could be explored to more directly encode molecular connectivity and physical laws, potentially enhancing interpretability and predictive accuracy for complex organic liquids.

5 Acknowledgements

We thank Oliver R. Gittus for his helpful guidance throughout this project.

6 References

- [1] Botejara-Antúnez, M., González-Domínguez, J., Rebollo-Castillo, F.J. and García-Sanz-Calcedo, J., 2025. Life cycle and environmental impact assessment of heat transfer fluids in parabolic trough CSP plants. *Sustainable Energy Technologies and Assessments*, 74
- [2] Martínez-Esteban, R.P., 2025. Environmental and health impacts of heat transfer fluids. *Science of the Total Environment*, 896
- [3] Wright, C.I., 2025. Enhancing the performance and durability of high-temperature heat transfer fluids in industrial applications: A short review. *Energy Storage and Conversion*, 3(1), Article 1959.
- [4] Electric Power Research Institute (2024). *Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption*. Report No. 3002028905. Palo Alto, CA: EPRI.
- [5] AXA XL Risk Consulting (2020) *Organic & Synthetic Heat Transfer Fluids and Equipment – PRC-7.1.5 Guidelines*. Available at: <https://axaxl.com/-/media/axaxl/files/pdfs/prc-guidelines/prc-7/prc715organicheattransferfluidsandequipmentv1.pdf>
- [6] Perry, R.H., Green, D.W. & Maloney, J.O. (eds.) (2008) *Perry's Chemical Engineers' Handbook*. 8th edn. New York: McGraw-Hill.
- [7] Lienhard, J.H., 2011. *A heat transfer textbook*. 4th ed. Cambridge, Mass.: Phlogiston Press.
- [8] Gharagheizi, F., 2013. Development of a general model for determination of thermal conductivity of liquid chemical compounds at atmospheric pressure. *AIChE Journal*, 59(9), pp.3322–3335
- [9] Bridgman, P.W., 1946. *General survey of certain results in the field of high-pressure physics*. Nobel Lecture, 11 December 1946.
- [10] Kwon, S.Y. & Lee, S.Y., 2012. Precise measurement of thermal conductivity of liquid over a wide temperature range using a transient hot-wire technique by uncertainty analysis. *Thermochimica Acta*, 542, pp. 18–23.
- [11] Assael, M. J., Antoniadis, K. D., Velliadou, D. & Wakeham, W. A., 2023. *Correct Use of the Transient Hot-Wire Technique for Thermal Conductivity Measurements on Fluids*. *International Journal of Thermophysics*, 44(6), p.85
- [12] Sastri, S.R.S. and Rao, K.K., 1999. A new temperature-thermal conductivity relationship similar to the Rackett equation for saturated liquid densities. *Chemical Engineering Journal*, 74, pp.161–169.
- [13] Chaparro, G. and Müller, E.A., 2024. Simulation and data-driven modeling of the transport properties of the Mie fluid. *Journal of Physical Chemistry B*, 128(2), pp.551–566
- [14] Braun, G., Nichele, J., Duarte, J.C., Alves, L.S.B. and Borges, I., 2025. Interpretable machine-learning models for predicting shear viscosity and thermal conductivity of binary fluid mixtures. *ACS Engineering Au*, 5(5), pp.573–583
- [15] Sargam, Y., Wang, K. and Cho, I.H., 2021. Machine learning based prediction model for thermal conductivity of concrete. *Journal of Building Engineering*, 34
- [16] Weininger, D. (1988) 'SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules', *Journal of Chemical Information and Computer Sciences*, 28(1), pp. 31–36.
- [17] Willett, P. (1998) 'Chemical similarity searching', *Journal of Chemical Information and Computer Sciences*, 38(6), pp. 983–996
- [18] Rogers, D. and Hahn, M. (2010) 'Extended-connectivity fingerprints', *Journal of Chemical Information and Modeling*, 50(5), pp. 742–754.
- [19] Lundberg, S.M., Erion, G.G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S.-I. (2020) 'From local explanations to global understanding with explainable AI for trees', *Nature Machine Intelligence*, 2, pp. 56–67.
- [20] Chapman, S. and Cowling, T.G. (1970) *The Mathematical Theory of Non-Uniform Gases*. 3rd edn. Cambridge University Press.
- [21] Hirschfelder, J.O., Curtiss, C.F. and Bird, R.B. (1954) *Molecular Theory of Gases and Liquids*. Wiley.

Optimising electrochemical biomass waste upcycling with automated experiments

Amiira Abdullahi and Shivani Parekh

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

The current dependence on fossil-fuel based chemical production has been hugely problematic due to its associated greenhouse gas emissions, particularly during the global climate crisis. Due to this, the electrochemical conversion of cheap feedstock like waste glycerol from the biodiesel industry into the more financially lucrative lactic acid has attracted a lot of interest. This work aimed to maximise the yield of lactic acid by varying flowrates, temperatures, current densities, and KOH concentrations and implementing Bayesian optimisation in an automated high-throughput setup to reduce the number of experiments required and human error, while also gathering more data in a shorter time frame. The catalyst in use was relatively inactive, thus the yields obtained were significantly lower than what has been seen in literature, however, the system was overall able to optimise this and would likely work to optimise the yields obtained from a more active catalyst. The trends of the changing parameters on the yield fit what was reported in literature, with this work giving additional insight into the differences across chronopotentiometry and chronoamperometry, and the benefits of the latter less harsh electrochemical method on improving yields and mitigating degradation.

1. Introduction

The electrochemical conversion of low-cost feedstocks into value-added chemicals has become an increasingly relevant research area, driven by the need to reduce high greenhouse gas emissions released from global dependence on fossil-fuel based chemical production.⁽¹⁾ One reaction that has gained particular interest is the electrochemical oxidation of glycerol. Around 90% of global glycerol production arises as a waste by-product of biodiesel manufacturing (10% w/w, with a surplus of 4.66 Mt), making it an abundant and inexpensive feedstock for electrochemical upgrading.⁽²⁾ Glycerol oxidation yields a variety of valuable products, including formic acid, glycolic acid, glyceric acid and lactic acid, which have significant demand across pharmaceutical, food and polymer industries. Lactic acid is particularly economically attractive due to its growing market value, projected to exceed 6.18 billion USD by 2034.⁽³⁾ However, over 90% of global lactic acid production currently relies on microbial fermentation, a process limited by long lag phases, strict operating conditions, contamination risks and costly downstream purification, making industrial production energy intensive and economically unfavourable.⁽⁴⁾

These challenges, combined with the abundance of waste glycerol, have led to significant interest in developing electrochemical routes for converting glycerol to lactic acid. Yet achieving high selectivity and yield remains difficult since commonly used catalysts such as Pt/C favour competing oxidation pathways,⁽⁵⁾ and lactic acid formation is highly sensitive to reaction conditions.⁽⁶⁾

The strong dependence on operating conditions, and the low yields and poor selectivity commonly reported for lactic acid, shows a clear need for systematic optimisation to find optimal conditions and increase lactic acid product. Hence, this project aims to identify the conditions, in particular temperature, electrolyte flowrate, KOH concentration and current density, that maximise lactic acid formation via glycerol oxidation using Bayesian optimisation and an automated electrochemical setup both to reduce error and to maximise throughput.

Previous studies have provided important insights into the competing reaction pathways in glycerol oxidation. Figure 1 provides a set of proposed pathways.^{(5), (7)} Glycerol is first electrochemically oxidised to glyceraldehyde or dihydroxyacetone, both of which are isomers and can be interconverted, where the equilibrium shift to dihydroxyacetone is favoured in more basic conditions. Papers propose that under alkaline conditions, dihydroxyacetone can undergo what is known as a dehydration-rearrangement-rehydration reaction to produce lactic acid, forming a pyruvaldehyde intermediate along the way.⁽⁸⁾ The other undesired pathway involves oxidation and carbon-carbon splitting to form glyceric acid, tartronic acid, glycolic acid, and formic acid. This shows that the ideal experimental conditions will guide the glycerol oxidation reaction down the first dehydrogenation pathway, making pyruvaldehyde the key intermediate. Therefore, as the reaction continues to optimise, the conditions should deter from the second oxidation pathway that results in glyceric acid.

Although the main products formed during glycerol oxidation are well documented, the underlying reasons why certain pathways dominate

under specific conditions remain poorly understood. Several studies highlight that the reaction mechanism is still not fully established at a molecular level, with limited insight into how intermediates form, and interact with the catalyst surface or solvent environment. Incomplete reaction data further hinder knowledge on the structural identity and behaviour of key species during conversion.⁽⁹⁾ This lack of mechanistic clarity makes it difficult to explain why some conditions favour oxidation to C₂–C₁ acids such as glycolic, oxalic and formic acid, while others promote retention of the C₃ backbone.⁽¹⁰⁾ Although temperature, pH and oxygen availability are known to influence product distribution, detailed explanations of how these specific parameters alter the reaction pathways have not been reported.⁽¹⁰⁾

Several studies have explored how reaction conditions influence which pathway dominates. As seen in Aslam, 2023, parameters such as temperature, electrolyte flowrate, KOH concentration and current density strongly affect catalyst–glycerol interactions and therefore the distribution of oxidation products.⁽¹¹⁾ In their work, the following variables were specified to remain constant; glycerol concentration in the anolyte, anion-exchange membrane, and the mass of Pt/C and Nafion in the electrode preparation. Zeolites were also employed in the construction of the catalyst to determine the effect of this on the lactic acid yield. While these studies provide useful information, due to the altering of the catalyst, it is unclear whether changes in yield were due to this or the parameters. It would be more apparent what parameters most strongly govern the selectivity toward lactic acid when the catalyst is kept constant. So, the quantitative relationship between key operating parameters established in literature (KOH concentration, temperature, current density, flow rate) and lactic-acid production when the catalyst is held constant is not fully understood and needs to be explored.⁽¹¹⁾ It is expected that higher temperatures, flowrates, and KOH concentrations would provide higher lactic acid yields due to the

associated increased reaction rates and reduced mass transfer limitations.⁽¹¹⁾

A review of previous literature revealed the effectiveness of Bayesian optimisation principles in solving complex and high-cost problems such as determining optimum reaction parameters.^{(12), (13)} For an objective function such as lactic acid yield, Bayesian optimisation enables the iterative balancing of exploration and exploitation by utilising appropriate acquisition functions resulting in locating global optima in the multivariate space. This is a much more evolved and efficient optimisation method in comparison to trial-and-error, one-factor-at-a-time (OFAT), and Design of Experiments (DoE) methods.⁽¹³⁾ Trial-and-error proved to be inefficient for multi-parameter reactions and thus did not determine global optima effectively, OFAT neglected to consider the interactions between variables, and although DoE methods via algorithms such as SNOBFIT accounted for relationships between variables and was better at determining global optima, it generally required significant data, thus raising costs.⁽¹²⁾ Despite its advantages, Bayesian optimisation has not been applied to the optimisation of glycerol electrooxidation for lactic acid production, ignoring a significant opportunity for methodological improvement.

This project builds on past literature but will aim to address the lack of clarity on how reaction parameters impact glycerol oxidation independent of the catalyst. This will be achieved by systematically mapping the parameters's influence on yield, selectivity and conversion. It will also improve on previous optimisation methods by introducing Bayesian optimisation as well as a novel automated high throughput set up.

2. Methods

2.1 Materials and chemicals

60 wt.% of platinum on carbon powder and Vulcan carbon black powder were sourced from ThermoFisher (3.6 nm Pt with trace metals). Glycerol (> 99.5%), potassium hydroxide

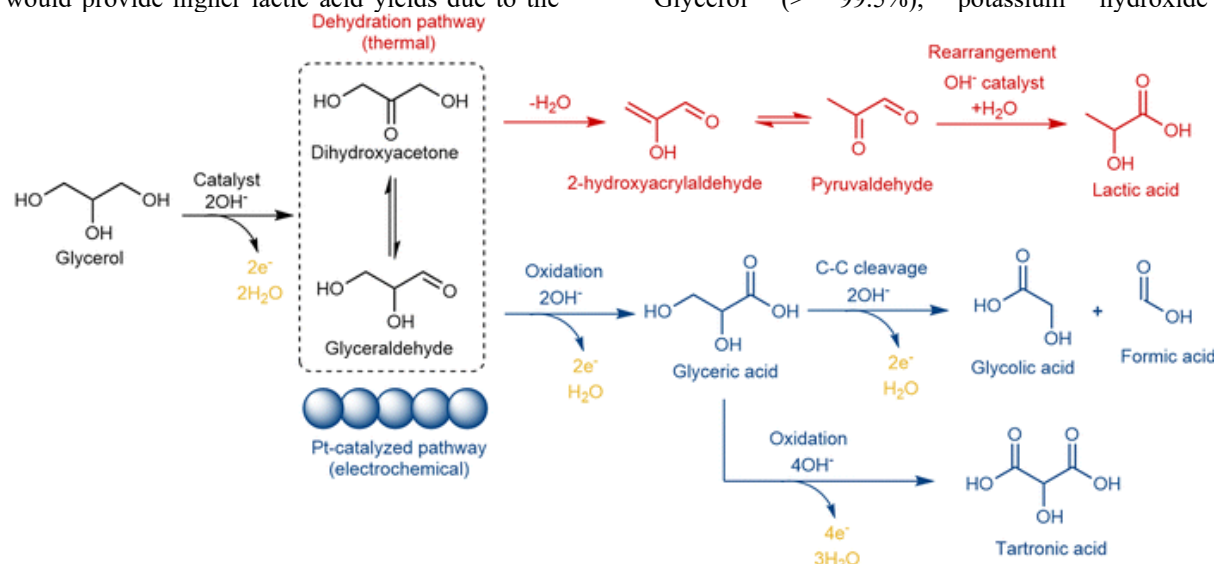


Figure 1. Glycerol oxidation reaction pathways⁽⁶⁾

(99.99%), ethanol absolute ($\geq 99.8\%$) and 99.9% isopropanol were obtained from Sigma Aldrich. Sulfuric acid (96%) was obtained from Supelco and Nafion® (D-521 dispersion, 5% w/w in deionised water and 1-propanol) from FUJIFILM Wako Chemicals. The anion exchange membrane was Fumasep FAA 3-50 and carbon paper was Freudenberg H23. A pre-loaded $0.5 \text{ mg}_{\text{Pt}} \text{ cm}^{-2}$ carbon paper (60wt% Pt/C on Sigracet 22 BB) was sourced from Fuel Cell Store.

2.2 Electrocatalyst preparation

The catalyst ink for the anode was made by combining 13.5 mg of 60 wt% Pt/C, 40.5 mg of carbon black, 16.8 mL of DI water, 6.34 mL of isopropyl alcohol, sonicating via probe sonication for 5 minutes, then for another 20 minutes after adding 0.864 mL of 5 wt% Nafion solution. The sonicated catalyst ink was spray coated on a carbon paper until a loading of $0.1 \text{ mg}_{\text{Pt}}/\text{cm}^{-2}$ was achieved. Spray-coating was automated with Cheersonic UAM4000L, employing a Pt-channel to avoid contamination. The catalyst-loaded carbon paper was cut into 2.5 cm x 2.5 cm squares for the anode and 3 cm x 3 cm for cathode.

2.3 High throughput flow cell setup

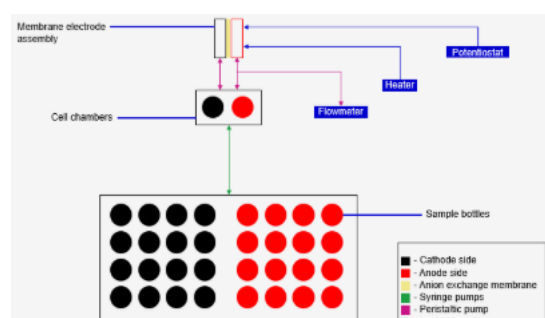


Figure 2. High throughput flow cell setup diagram

The high throughput flow cell setup was aimed to automate up to 16 electrochemical reactions. This increases the amount of data that can be collected in one experimental session, as compared to running each reaction separately. The setup was built in-house and remains in the development stage.

The catalysts for the anode and the cathode alongside the anion exchange membrane were pressed into the membrane electrode assembly (MEA) and fixed into place by bolting

the cell shut.

As shown in Figure 2, the syringe pumps can move the relevant samples into the chamber for testing and then return them to the original bottle. 10 mL of the desired initial concentrations of KOH and KOH + 1M Glycerol was loaded into the cathode and anode bottles respectively. Tubing was fixed onto each sample bottle so that syringe pumps could draw up the unreacted solutions into the chamber bottles from which the solutions were run through the cell under an applied potential (chronoamperometry) or current (chronopotentiometry). A python script was employed to control the potentiostat for the electrochemical measurements, the flowrates via a peristaltic pump, and the temperature using a CAL 3200 proportional-integral-derivative (PID) temperature controller with feedback from a thermocouple at the anode. A Bronkhorst M13 mini CORI-FLOW MFM flowmeter connected to Bronkhorst FlowSuite software was also integrated in the python to automatically measure the flowrate across experiments.

There were up to 16 reactions that each proceeded consecutively so multiple samples could be run over the same catalyst, and more data could be collected. Each sample ran for around one hour, this fixed time was to allow for products to form but not for too long so that the impact of the other parameters can be seen and compared. The reacted solutions would then be pumped back into the chambers and shortly after collected back into the sample bottles. Between the execution of each trial, the system was thoroughly flushed with deionised water for 15 minutes using the peristaltic and syringe pumps

2.4 Bayesian optimisation

Bayesian optimisation was implemented within the experimental process by utilising a NIMO python library sourced via GITHUB. The objective function was set to be the yield of lactic acid, and the variable parameters to be KOH concentration, current density, flowrate, and temperature. The values that each variable can take can be seen in Table 1.

A simplified schematic of the optimisation process can be seen in Figure 3. Four key python files were employed in this process: the first to

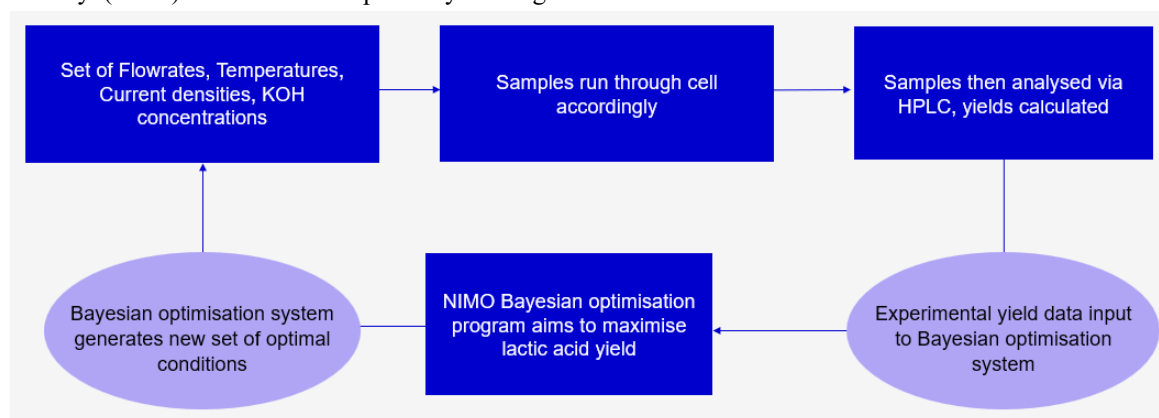


Figure 3. Simplified Bayesian optimisation flowsheet

generate an excel file of every combination of the reaction conditions (the candidates), the second to randomly select 11 of the candidates (initial proposals), the third to actually run the cell set-up based on the input proposal conditions.

From these run conditions, the reacted samples were analysed to calculate the lactic acid yield, which was then input as objective function values in the candidates file. Finally, the fourth python file was run to process the yield data input into the candidates file and generate a new set of optimised proposals. This process was repeated seven times to find the optimum reaction conditions within the designed search space

Table 1. Reaction parameters

[KOH] (M)	Flowrate (rpm)	Current density (mA cm ⁻²)	Temperature (°C)
0.5, 1, 2, 3	10, 20, 30 40, 50	4, 10, 20, 24, 30, 36	25, 30, 40, 50, 60

2.5 Product analysis

High Performance Liquid Chromatography (HPLC) was employed through Agilent's OpenLab ChemStation with a Refractive Index Detector (RID) and a Variable Wavelength Detector (VWD). The Aminex column was used with a flowrate of 0.4 ml min⁻¹ and a mobile phase of 5mM H₂SO₄.

Each sample that needed to be analysed was first neutralised by an equal volume of aqueous sulfuric acid, the concentration of which was determined based on the concentration of KOH in the sample as seen in Table 2. The neutralised samples were next loaded into the ChemStation machine, and the sequence was run under an RID temperature of 50°C. The peaks from the analysed samples were then integrated by RID and VWD detectors, as seen in Figure 4, to calculate the concentration of each product. In cases where auto-detection failed, the peaks were manually integrated and split, especially in the case of lactic acid and glycolic acid, as seen in Figure 5, introducing error in the analysis. The concentrations were then used in equations 1, 2, and 3 to calculate the yield, selectivity, and conversion. Equation 2 was used as the percentage yield calculation, as the theoretical moles of lactic acid produced is the same as the actual moles of glycerol reacted due to the stoichiometric ratio of the reaction being 1:1.

Table 2. Concentrations of KOH and the associated concentration of sulfuric acid needed to neutralise the HPLC sample

[KOH] (M)	0.5	1	2	3
[H ₂ SO ₄] (M)	0.3	0.55	1.05	1.55

$$\text{Conversion} = \frac{\text{mol of glycerol reacted}}{\text{initial mol of glycerol}} \quad (1)$$

$$\text{Yield} = \frac{\text{mol of lactic acid produced}}{\text{mol of glycerol reacted}} \quad (2)$$

$$\text{Selectivity} = \frac{\text{mol of lactic acid produced}}{\text{total mol of products}} \quad (3)$$

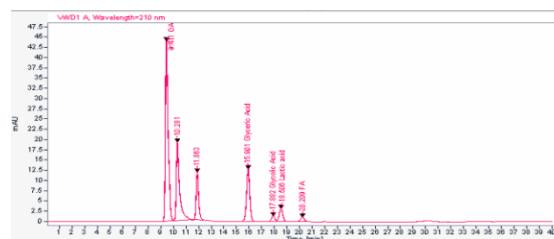


Figure 4. Products seen on the HPLC under VWD

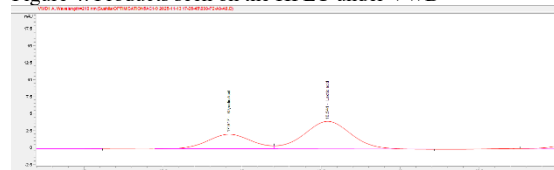


Figure 5. Manually split lactic acid and glycolic acid peaks

2.6 Electrochemical measurements

An Ivium-n-stat potentiostat, controlled by a python script, was used to conduct electrochemical experiments. Diagnostic measurements were employed in the form of cyclic voltammetry (CV) and open circuit voltage (OCV) to ensure that the fresh catalyst in the MEA was active, and the assembly of the MEA was satisfactory.

For CV the potential applied was scanned between 0 and 1.1V at a rate of 50 mV/s for 5 cycles, while measuring the current response.

The OCV was measured across the cell for 5 minutes.

The catalyst was activated before each run by applying a current of 4 mA across the MEA and the response potential was measured for a period of 20 minutes. This process was implemented to maximise the catalyst activity by unblocking the active sites and increase wetting of the catalyst by releasing microbubbles of air limiting mass transfer.

For chronopotentiometry (CP) measurements, the desired current was applied to the working electrode (anode) for 1 hour, and the potential was measured. The current was held constant between the anode and the counter electrode (cathode), and the anodic response potential was measured relative to the reference electrode connected to the cathode where hydrogen evolution reaction (HER) occurs.

For chronoamperometry (CA) measurements, the desired potential was applied to the anode, and the current was measured. A reversible hydrogen electrode (RHE) was placed in the anolyte chamber for chronopotentiometry experiments to stabilise the applied potential.

For both the CA and the CP measurements, the current density was calculated according to the area of the anode (2.5 cm x 2.5 cm).

3. Results and discussion

3.1 Reproducibility run

A reproducibility run was conducted over the MEA, where the standard conditions of 1M KOH, 60 °C, 50 rpm, and 20 mA cm⁻² were run 11 times over the same catalyst. The parameters used were selected from the optimal glycerol oxidation reaction conditions determined in Aslam, 2023.⁽¹¹⁾ This was done to determine if results were reproducible from the literature as well as to define a benchmark for how many runs can be completed on the same catalyst before degradation.

Figure 6 provides a breakdown of the conversion, yield, and selectivity for each run. The yield and conversion were observed to be reasonably consistent across the runs after the first three samples. This inconsistency may be due to a technical fault that induced a time delay between the running and analysis of the first three run and the next eight. During this time delay, the catalyst surface may have changed, leading to inconsistency in the reactions between samples 1-3 and 4-11. Overall, the yield remained below 5 %, significantly lower than the lowest yield of 16.8 % as seen for the same conditions in Aslam, 2023.⁽¹¹⁾ Additionally, as seen in Figure 7, a consistent amount of lactic acid was produced, however, varying amounts of formic acid, glycolic acid, and glyceric acid were produced as side products. The standard deviation in the selectivity remained at 4.13 % with the first three runs excluded, exhibiting relative inconsistency in selectivity for this reproducibility run in comparison to yield. However, overall, the results appear to be reliable.

Reasons for this minimal variance in the repeated standard runs could include human error in carrying out the experiments as this was done with less prior experience in comparison to later runs, thus introducing variation in these controlled experiments. Additionally, an explanation for the diminished yield could be surface fouling due to the adsorption of intermediates blocking active sites of the catalyst, resulting in the lactic acid pathway not being favoured.

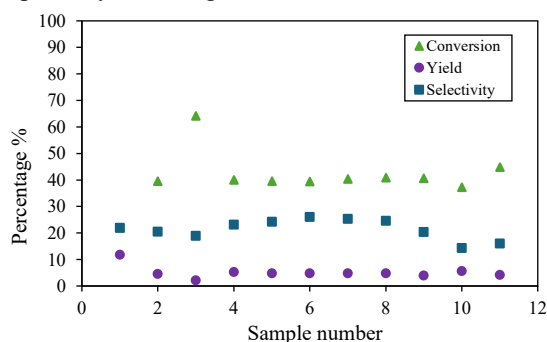


Figure 6. Conversion, yield, and selectivity of each reproducibility sample

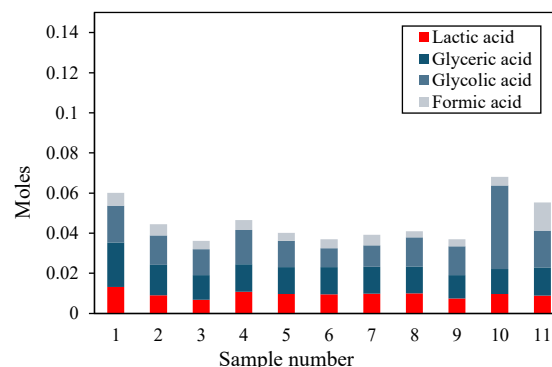


Figure 7. Product breakdown of each sample in moles

3.2 Single-objective optimisation runs (1-4) using chronopotentiometry

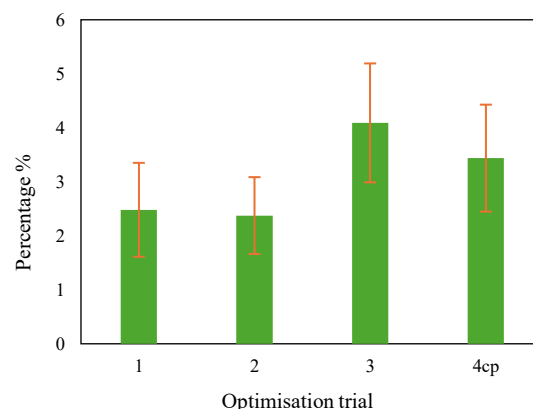


Figure 8. Averaged yield of each trial

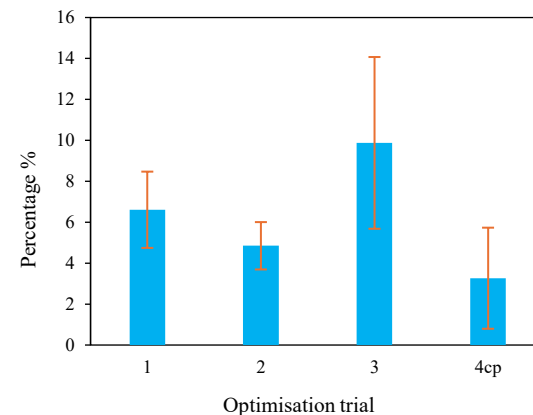


Figure 9. Averaged selectivity of each trial

The average conversion across the samples of glycerol remained at approximately 28 % for all four optimisation runs, all completed via chronopotentiometry, thus a similar amount of glycerol is consistently reacting. Hence, when a larger amount of lactic acid is produced causing a larger yield, the reaction favours the lactic acid pathway increasing selectivity, therefore the yield and the selectivity trends should mirror one another. Additionally, as calculated by Equations 2 and 3, the selectivity would be higher than the yield due to products such as tartronic acid being unidentified by HPLC due to lack of calibration, thus the total moles of products being less than the

total moles of glycerol reacted. This is seen in Figures 8 and 9, where the average yield and selectivity appeared to follow the same trend, where the largest increase can be seen for both from run 2 to run 3. Furthermore, there is a notable decrease from run 3 to run 4, and from run 1 to run 2.

However, the values of the yield remain significantly lower than Aslam, 2023, evidencing that these runs have not optimised the reaction conditions to maximise the lactic acid yield.⁽¹¹⁾ Additionally, whilst gathering data, it was noted that several of the reacted samples had a dark brown or dark yellow discolouration, indicating catalytic degradation.

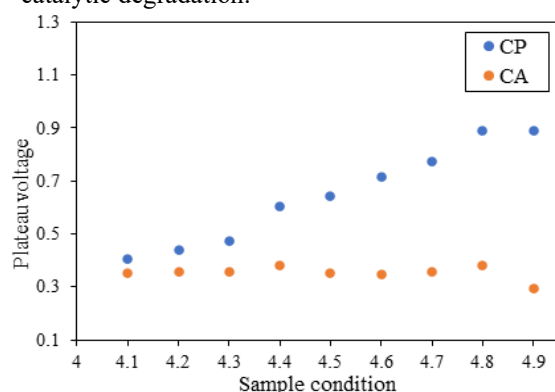


Figure 10. Comparison of plateau voltages from subsequent runs for both CP and CA, both performed under optimisation 4 proposal conditions

The trend that ‘4CP’, optimisation run 4 under chronopotentiometry, followed on Figure 10, establishes that the plateau values of potential for the activation runs continued to increase after each run. This increase in potential response to the same applied current indicates that the overpotential required to carry the fixed current of 4 mA has increased, possibly due to fewer available active sites on the catalyst, which reinforces the possibility of catalyst degradation or surface fouling. Such degradation may be responsible for reduced catalytic activity and poor lactic acid yields; thus, it was imperative to implement improvements to the experimental method.

3.3 Improving the experimental method

3.3.1 Diagnostic check

The first implemented method improvement was utilising diagnostic OCV and CV scans before a fresh catalyst was input to the MEA and including these curves between each run to verify a new reaction was not initiated on a degrading catalyst. This was the first line of quality control to identify issues in the cell setup, as this could cause immediate catalyst degradation if the potentiostat applied a high current. The aim was to adjust the setup of the MEA by changing the catalyst if signs of decay were evident between samples, or if there was poor catalyst-membrane contact due to poor

loading in the MEA, or to tighten the assembly in the case of leaks.

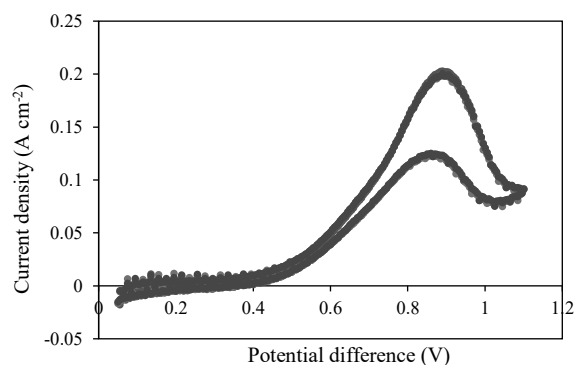


Figure 11. Example diagnostic cyclic voltammetry curve

For the diagnostic measurements on a fresh catalyst prior to reactions occurring, the standard concentrations of 1M KOH and 1M KOH + 1M glycerol were pumped into the cathode and the anode respectively, and the OCV and CV scans were performed.

As seen in Figure 11, forward scan showed a peak between 0.8 V and 1 V and a current density of 0.2 A cm⁻² in this case of a less active catalyst. This catalyst had not undergone any reactions, thus the lower current density could not be attributed to degradation, but perhaps to a lack of wetting of the catalyst which would decrease the contact of reactants with the catalyst surface, thus reducing the electrochemical surface area and catalytic activity. The strong anodic peak indicates multiple electron transfer steps in the oxidation of glycerol, while the weaker cathodic peak indicates that there could have been some surface fouling of the catalyst due to the strong adsorption of products.⁽¹⁴⁾ This fouling can reduce the number of active sites on the catalyst to adsorb, thus lower the activity of the catalyst and the subsequent lactic acid yields. This passivation layer over the catalyst can also lead to catalyst stability and reduce the possibility of degradation when the catalyst is subjected to higher pHs or current densities.

3.3.2 Chronoamperometry

Comparing the performance of the catalyst under chronoamperometry was successfully attempted for the same proposal reaction conditions as optimisation run 4 under a chronopotentiometry run. This was implemented after noting that the activation curves generated before each chronopotentiometry run plateaued at an increasingly higher potential after each run, as seen on Figure 10, indicating catalyst degradation. This indicated that the catalyst continued to degrade after each reaction, due to the system driving the potential to increasingly higher values as it was being subjected to higher current densities than the surface could sustain. This in turn did not promote the glycerol oxidation reaction. To implement this change, the current values from the optimisation run 4 proposals were exchanged for potential

values as seen in Table 3. These values were selected as the associated held current would induce that particular response potential, in addition to being in the range of potentials for which the catalyst would be active for glycerol oxidation. From Figure 10, it can be noted that there is reduced catalyst deactivation after each reaction. This can be attributed to holding the potential at a fixed value allowing the flow of charge to change corresponding to the catalyst surface condition, reducing degradation.

Table 3. The held potential values for chronoamperometry corresponding to the held current values in chronopotentiometry

CP (mA)	4	10	20	24	30	26
CA (V)	0.4	0.5	0.6	0.7	0.8	0.9

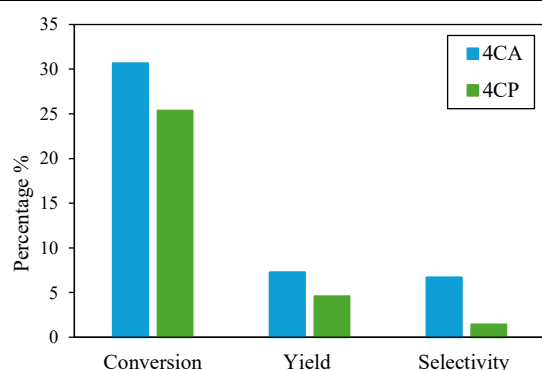


Figure 12. Comparison of conversion, yield, and selectivity for the same run conditions subject to CP and CA

Additionally, it was noted that the conversion, yield, and selectivity improved when switching to CA over CP, at an average percentage increase of 14.5%, 20.6%, and 148.1% respectively. Correspondingly, from Figure 12, the increase in all three metrics for a randomly selected condition run under both CA and CP can be seen. These outcomes can be attributed to the decreased degradation of the catalyst under CA.

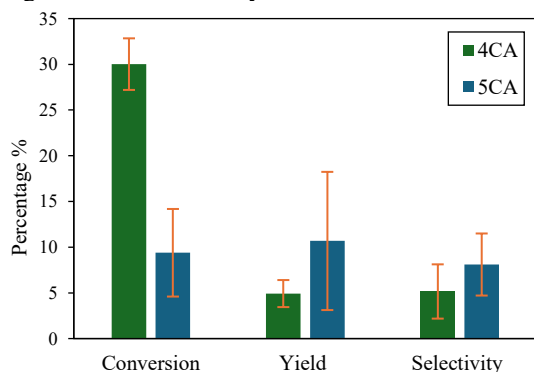


Figure 13. Conversion, yield, and selectivity comparison of runs 4 and 5

Due to these positive results, optimisation run 5 was completed under CA, and as seen from Figure 13, there was a significant increase in yield and selectivity from run 4 under CA, a percentage increase of 117% and 57% respectively.

3.3.3 Multi-objective optimisation

Another method improvement that was employed to improve the performance of the Bayesian optimisation system was implementing multi-objective optimisation in the program. The addition of an objective function such as selectivity could work to increase yield by promoting the lactic acid pathway over side reactions, ensuring that a larger amount of lactic acid is produced.

This process was implemented to enhance the improvement noted from optimisation runs 4 to 5, both of which were performed under CA. The yield and selectivity data from both of these runs was input to the multi-objective program to generate a new set of optimisation run 6 proposals.

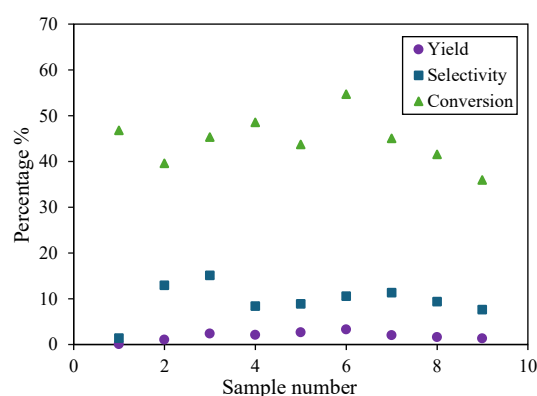


Figure 14. Conversion, yield and selectivity for each run under multi-objective optimisation

However, as noted from Figure 14, there was not a significant improvement in any of the metrics. This can be attributed to the similarity between these two objectives leading to negligible improvement in results when implementing both.

3.3.4 Fresh catalyst

The final method change that was performed was spraying an entirely new catalyst. This was implemented to try to maximise yield further as we had optimised as much as we could for the original catalyst. For this new catalyst, single objective optimisation was implemented once again; to test the amount of degradation this freshly sprayed catalyst would experience under standard conditions.

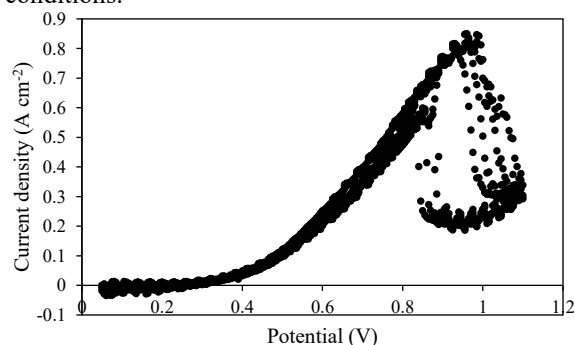


Figure 15. Diagnostic cyclic voltammetry curve for the new catalyst

From this, it was noted that the CV diagnostic curves had improved in terms of the peak currents associated with higher catalytic activity, as seen in Figure 15.

Higher yields of 20 % under standard conditions were also reached, which were closer to literature values. This new catalyst performed significantly well relative to the previous results of each of the setup changes that were included, supporting prior

literature that the catalyst itself is the most influential change.⁽¹¹⁾ This also demonstrated the limitations in finding the optimal conditions for lactic acid yield if the catalyst has poor activity as this hindered the ability of the system to reach yields comparable to those in literature.

3.4 Heat maps

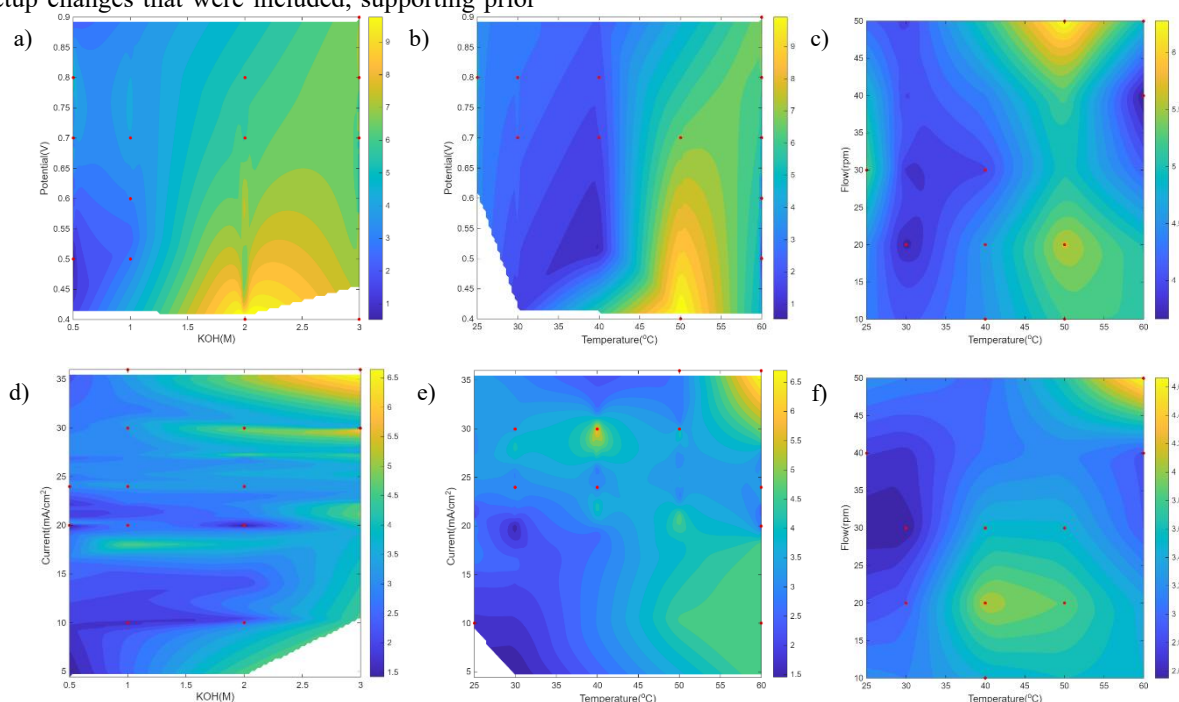


Figure 16. Heat maps for yield vs parameters. (a) Potential vs KOH for CA (b) Potential vs temperature for CA (c) Flow vs temperature for CA (d) Current vs KOH for CV (e) Current vs Temperature for CV (f) Flow vs Temperature for CV

To visualise the effects of the different reaction parameters on lactic acid yield, heatmaps were plotted based on NIMO Bayesian optimisation predictions calculated from experimental data. Figure 16 displays 2D heat maps to demonstrate the variation of the objective function, where yellow areas indicate values of high yield and blue areas for low yield values. Graphs 16b) and 16e) show the changes in yield for applied potential against temperature and applied current against temperature for CA and CP data respectively. Graphs 16a) and 16d) compare the variation in yield for applied potential against KOH concentration and applied current KOH concentration for CA and CP data respectively. Graphs 16c) and 16f) compare flowrate against temperature for both CA and CP data. The model generated predictions by leveraging the experimental yield data that was input, conveyed by the red points, and producing yield prediction data for all of the unexplored points within the defined search space. These heat maps act as a useful method to analyse the trends in yield of these combinations of parameters.

Graphs 16a) and 16d) indicate that higher concentrations of KOH maximise the yield of lactic

acid consistently across all of the figures and persists for both CA and CP. For CA, higher yields were observed for lower KOH concentrations, as at higher pH, the carbon-carbon bond cleavage to form side products such as formic acid and glycolic acid, thus in overly basic environments, there is the possibility of reduced yield and selectivity. Contrastingly, for CP, higher concentrations of KOH were required due to achieve higher yields. This trend was supported by literature as more basic conditions were correlated with shifting the glyceraldehyde equilibrium reaction to produce more dihydroxyacetone, which would further react to produce lactic acid.⁽¹¹⁾ The lower required KOH concentration for CA than CP indicates that employing CA could reduce the operating costs associated with this electrochemical process.

From graphs 16b), 16c), 16e) and 16f), it is notable that CA generated higher yields for relatively lower temperatures of 50°C and potentials, while CP promoted higher current densities and relatively higher temperatures of 60°C, for yields that yet fell short of those from CA. An explanation for this could be the degradation of the catalyst or surface fouling caused by CP reducing the number of available

active sites on the catalyst for the reaction to occur, thus higher temperatures being promoted. Higher temperatures have been noted to improve the overall glycerol oxidation reaction rate according to literature, by increasing the diffusion of the reactants to the surface, improving the adsorption of key intermediates in the lactic acid pathway, and by increasing the energy level of the reactants which also increases the rate of the dehydration step.⁽¹¹⁾ Additionally, low held potentials were promoted for CA in comparison to high held currents being promoted for CP. This further reinforces the notion that CA does not reduce the electrochemical surface area of the catalyst by blocking the active sites, ensuring that lower potentials can induce oxidation of glycerol.

The effect of a larger flowrate as seen from Graphs 16c) and 16f) is significantly more beneficial for the yield for both CP and CA due to the existent catalyst degradation causing fewer active sites for adsorption of key lactic acid intermediates. Increasing the flowrate of the analyte enables the system to overcome the mass transport limitations of the reactant to the catalyst surface, thus increasing the rate of reaction. Notably, the values of the yield achieved by CA were still higher than CP.

4. Conclusions

In conclusion, this work enabled the exploration of the impact of temperature, flowrate, current, and concentration on the lactic acid yield for a glycerol oxidation reaction. The novel automated high throughput setup supported the generation and processing of a large data set over a short period of time and reduced associated human error significantly.

The use of Bayesian optimisation facilitated the exploration of all four key variables with less experiments than would otherwise be needed; however, although optimal yield conditions were found, the yields were significantly lower than those found in literature due to key problems with the catalyst in use. It was also confirmed that in the case of non-conflicting metrics such as selectivity and yield, implementing multi-objective Bayesian optimisation has a negligible effect.

Introducing cyclic voltammetry and open circuit voltage diagnostic steps was also valuable in evaluating possible issues with the catalyst. These techniques should be implemented particularly in all high throughput setups, particularly as a measure to examine whether catalyst deactivation is inducing low yields in some of the data, even in the case of a more active catalyst than the Pt/C benchmark used in this work.

Another troubleshooting step that provided valuable insight was the implementation of chronoamperometry. Alongside the yields from this electrochemical method improving to around 10% for an already relatively inactive catalyst, the

reduced degradation and surface fouling was noted to make it a gentler method in comparison to chronopotentiometry.

Additionally, the trends noted for chronopotentiometry in the heat maps provided matched what was expected from literature in terms of increased yield being seen in the cases of higher temperature, flowrate, and KOH concentration. The added insight from chronoamperometry also enabled understanding of how the performance of a Pt/C catalyst less prone to degradation is affected by the reaction conditions.

However, when a freshly sprayed catalyst was employed, a significantly larger yield was achieved. This reinforces the idea that the catalyst itself is hugely influential to the yields that are able to be reached, but despite this, the setup was still able to optimise the yields for this deactivated catalyst.

5. Future works

For future consideration, it would be useful to implement further spectroscopic diagnostic tests surrounding the catalyst degradation. This could be in the form of implementing Raman spectroscopy to track changes in the crystal structure of the carbon and platinum and indicate if deactivation is due to surface fouling.

In addition to this, a stable operating range for the chosen catalyst could be identified by first altering one parameter at a time prior to the experimental runs to ensure that the parameter ranges are not harsh enough to induce degradation. Moreover, continuous variables could be introduced to increase the search space of the optimisation program by enabling the exploration of more conditions.

Finally, a more stable and active catalyst, like Pt/Au or Pt/Pd could be employed to measure the hopefully more significant increase in yield that the optimisation programme is capable of inducing.

6. Acknowledgements

The authors would like to thank Dr Sushila M Marlow for her continued support in this project, alongside everyone in the Petit, Titirici, and Pini groups at Imperial College London for their encouragement and guidance.

7. References.

1. *Electrochemical Conversion of Biomass Derived Products into High-Value Chemicals*. **P. Prabhu, Yan Wan, Jong-Min Lee**. 2020, CellPress, pp. 1162-1177.
2. *Environmental impacts of valorisation of crude glycerol from biodiesel production – A life cycle perspective*. **Marco Tomatis, Harish Kumar Jeswani, Adisa Azapagic**. 55-65, Manchester : Elsevier, 2024, Vol. 179.
<https://doi.org/10.1016/j.wasman.2024.03.005>.
3. *Metabolic engineering as a tool for enhanced lactic acid production*. **Bikram P Upadhyaya, Linda C. DeVeaux, Lew P. Christopher**. 12,

- South Dakota : CellPress, 2014, Vol. 32.
<https://doi.org/10.1016/j.tibtech.2014.10.005>.
4. *Lactic Acid: A Comprehensive Review of Production to Purification*. **Abidemi Oluranti Ojo, Olga de Smidt**. 688, Bloemfontein : Processes, 2023, Vol. 11.
<https://doi.org/10.3390/pr11030688>.
5. *Selective glycerol to lactic acid conversion via a tandem effect between platinum and metal oxides with abundant acid groups*. **Hui Luo, Mianle Xu, Sihang Liu, Giulia Tarantino, Hanzhi Ye, Hossein Yadegari, Alain Y. Li, Ceri Hammond, Georg Kastlunger, Ifan E. L. Stephens, Maria-Magdalena Titirici**. 87-96, London : Royal Society of Chemistry, 2024, Vol. 3.
<https://doi.org/10.1039/D4EY00236A>.
6. *Progress and Challenges in Electrochemical Glycerol Oxidation: The Importance of Benchmark Methods and Protocols*. **Hanzhi Ye, Silvia Favero, Helen Tyrrell, Kanyapat Plub-in, Anna Hankin, Reshma R Rao, Ifan E. L. Stephens, Maria-Magdalena Titirici, Hui Luo**. 13, London : ChemCatChem, 2025, Vol. 17.
<https://doi.org/10.1002/cctc.202500152>.
7. *An efficient one pot conversion of glycerol to lactic acid using bimetallic gold-platinum catalysts on a nanocrystalline CeO₂ support*. **Rajeesh Kumar Pazhavelikkakath Purushothaman, J. van Haveren, D.S. van Es, I. Melián-Cabrera, J.D. Meeldijk, H.J. Heeres**. 92-100, Groningen : Elsevier, 2013, Vol. 147.
<https://doi.org/10.1016/j.apcatb.2013.07.068>.
8. *THE MECHANISM OF CARBOHYDRATE OXIDATION. XII. THE ACTION OF POTASSIUM HYDROXIDE ON l-ARABINOSE AND d-XYLOSE*. **Wm. Lloyd Evans, Rollin Francis Conaway**. 9, 1930, Vol. 52.
<https://doi.org/10.1021/ja01372a037>.
9. *Unveiling the Mechanism of Glycerol Oxidation to Lactic Acid on Pt/Sn-MFI Zeolite: an In situ Solid-state NMR Study*. **Shen Xueyuan, Qi Guodong, *, Liang Jiawei, Wang Ruichen**. 935–942, s.l. : Chemical Research in Chinese Universities, 2024, Vol. 40.
<https://doi.org/10.1007/s40242-024>.
10. *Enhancing the understanding of the glycerol to lactic acid reaction mechanism over AuPt/TiO₂ under alkaline conditions*. **Christopher D. Evans, Mark Douthwaite, James H. Carter, Samuel Pattisson, Simon A. Kondrat, Donald Bethell, David W. Knight, Stuart H. Taylor, Graham J. Hutchings**. 134705, s.l. : The Journal of Chemical Physics, 2020, Vol. 152.
<https://doi.org/10.1063/1.5128595>.
11. *Efficient and selective glycerol electrolysis for the co-production of lactic acid and hydrogen with multi-component Pt/C-zeolite catalyst*. **Muhammad Aslam, Miriam Navlani-García, Diego Cazorla-Amorós, Hui Luo**. 1, s.l. : JPhys Materials, 2023, Vol. 7. 10.1088/2515-7639/ad0561.
12. *Bayesian Optimization for Chemical Synthesis in the Era of Artificial Intelligence: Advances and Applications*. **Runqiu Shen, Guihua Luo, An Su**. 2687, s.l. : Processes, 2025, Vol. 13.
<https://doi.org/10.3390/pr13092687>.
13. *Cost-Aware Bayesian Optimization for Self-Driven Condition Screening of Flow Electrosynthesis*. **Runzhe Liang, Siyuan Zheng, Kai Wang, Zhihong Yuan**. 3, s.l. : ACS Electrochemistry, 2024, Vol. 1.
<https://doi.org/10.1021/acselectrochem.4c00078>.
14. *Spectroelectrochemical studies on the effect of cations in the alkaline glycerol oxidation reaction over carbon nanotube-supported Pd nanoparticles*. **Dennis Hiltrop, Steffen Cychy, Karina Elumeeva, Wolfgang Schuhmann, Martin Muhler**. 1428-1435, s.l. : Beilstein Journal of Organic Chemistry, 2018, Vol. 14.
<https://doi.org/10.3762/bjoc.14.120>.

Enhanced Fresh Water Desalination Using Barium Bismuth Oxide with Efficient Energy Management

Maya Yastrib, Gilby Chatha

Department of Chemical Engineering, Imperial College London, SW7 2AZ, United Kingdom

Abstract

With the current global water crisis and with less than 1% of the planet's water available as safe and accessible freshwater, finding sustainable and efficient methods to produce clean water is of great importance. Developing research on photothermal materials and interfacial solar evaporation shows promising results compared to the more conventional desalination technologies due to lower energy requirements and the use of renewable solar energy. This report focuses on the performance of barium bismuth oxide ($\text{Ba}_2\text{Bi}_2\text{O}_6$) as a photothermal material for freshwater production via desalination. The investigation first identifies the optimal conditions for $\text{Ba}_2\text{Bi}_2\text{O}_6$ synthesis of this material and then determines its ideal mass loading, before evaluating the effectiveness of a specially designed device made up of evaporation and condensation compartments aiming to simulate an open system. It was found that high-purity $\text{Ba}_2\text{Bi}_2\text{O}_6$ was produced after 90 minutes of high energy ball milling at 500 rpm, followed by calcination at 700°C for 2 hours in a tube furnace. The purity and optical properties of the synthesised material were characterised using X-ray diffraction and UV-Vis diffuse reflectance spectroscopy. After measuring temperature change, evaporation rate and efficiency, a mass of 50mg proved to be optimal for the loading. Desalination was then measured using a water sample and a seawater sample, with the percentages of clean water recovery on a mass basis equating to 1.59% and 4.55% respectively. Although promising, water collection can be increased by future modifications to device such as improving thermal insulation, optimising condensation surface as well as refining the $\text{Ba}_2\text{Bi}_2\text{O}_6$ loading mass.

1. Introduction

Access to clean and safe freshwater is a global challenge that is becoming increasingly concerning due to several factors including the expansion of industry sectors, population growth and a greater demand for clean water from industries such as agriculture. With only 0.5% of the earth's water safe for human consumption^[1], it is expected that the demand for this resource will outstrip supply by 40% by the end of the decade^[2], intensifying the need to find more sustainable production methods. With large-scale process like reverse-osmosis desalination and distillation being the main ways of water production, it is important to prioritise alternatives that are less energy intensive and have a lower carbon footprint than conventional methods driven by fossil fuels.

Recent developments in photothermal materials have significantly improved their solar photon capture abilities and photothermal conversion performance, opening up an expanding range of applications. Modern photothermal systems involve metallic and semiconductor nanostructures and carbon based, organic polymer and

two-dimensional (2D) nanomaterials. Through integration with other molecules or particles, the functionalised photothermal materials with synergetic properties can perform multiple tasks for targeted application. Therefore, photothermal materials are now being explored in diverse technologies such as photocatalysis, photothermal therapy and solar driven water evaporation. Although many of these concepts remain at the laboratory scale, several successful real-world implementations exist, such as the PCR test and AutoLase Therapy^[3]. To enable further commercialisation, critical considerations such as the ability to scale up with optimised cost efficiencies, biosecurity and environmental safety of the materials, are crucial. This motivates continued efforts to optimise existing systems and develop new materials.

Photothermal materials can be successfully used for interfacial solar evaporation (ISE), which is an emerging and promising solution to this worldwide crisis due to solely relying on solar energy to drive evaporation, resulting in minimal

energy input and lower operational costs. This photon-to-thermal conversion occurs through non-radiative recombination, as shown on *Figure 1.1A*. Non-radiative recombination occurs between excited electrons in the conduction band and trap states from defects and dopants, with the released energy being in the form of phonons (heat). This heat is localised at the air-water interface instead of the bulk water, resulting in more efficient evaporation. A condenser made of a suitable material is then used to collect the clean water. However, various challenges have arisen with this method, such as limited evaporation with large scale applications, material durability and integration with condensation systems. Therefore, it is vital to address these issues when transitioning from laboratory prototypes to larger scale systems.

The objective of this report focuses on the suitability of $\text{Ba}_2\text{Bi}_2\text{O}_6$ as a semiconductor photothermal material, which is a potential candidate for clean water production, as illustrated in *Figure 1.1B*, due to its double perovskite structure. Double perovskites have a broad solar absorption spectrum and a narrow bandgap allowing for greater heat generation as well as thermal stability, surviving longer under solar exposure. Additionally, $\text{Ba}_2\text{Bi}_2\text{O}_6$ possesses low thermal conductivity, facilitating heat localisation at the air-water interface, therefore enhancing interfacial evaporation as opposed to heating the bulk liquid. The material's nontoxicity and relatively low cost also encourage further research on whether this could be more feasible than alternative photothermal options. Although previous research has studied the effects of other photothermal materials such as black TiO_2 and MXenes, there are few investigations involving $\text{Ba}_2\text{Bi}_2\text{O}_6$ so comparing its performance and efficiency to existing materials is essential.

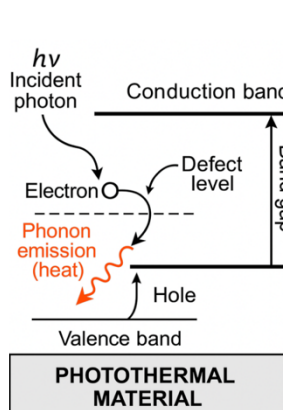


Figure 1.1A Schematic non-radiative recombination.

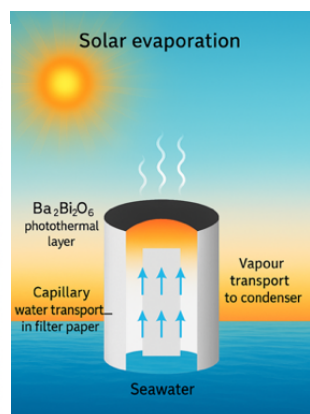


Figure 1.1B Schematic of interfacial solar evaporation using $\text{Ba}_2\text{Bi}_2\text{O}_6$.

This investigation consisted of measuring the temperature change and the evaporation rate of $\text{Ba}_2\text{Bi}_2\text{O}_6$ to find the optimum loading, collection rate in the novel condenser device and saline water evaporation stability. To collect these data, a device was used consisting of an evaporation unit with the filter paper containing $\text{Ba}_2\text{Bi}_2\text{O}_6$ and a condenser unit. This isolated container was also built to create and maintain a stable environment for these two processes as well as to keep the filter paper in place during water absorption. To improve the performance of this material, defects were introduced through high energy ball milling (HEBM) allowing for enhanced light absorption.

Following the production of $\text{Ba}_2\text{Bi}_2\text{O}_6$, various analytical processes were used to determine its effectiveness including XRD to determine purity and UV-Vis to measure the light absorption spectrum.

2. Methodology

2.1 Synthesis of Barium Bismuth Oxide

Barium Bismuth Oxide ($\text{Ba}_2\text{Bi}_2\text{O}_6$) was prepared via a solid-state route using barium acetate and bismuth acetate as precursors. Stoichiometric amounts of each material were measured to achieve a 1:1 molar ratio, corresponding to 0.225g of barium acetate and 0.386g of bismuth acetate. Acetates were selected based on their thermal decomposition that predominantly yields benign volatile by-products, such as water and carbon dioxide. The powders were mixed using high energy ball milling at 500 rpm for 90 minutes to ensure mechanical activation, control microstructure and partial amorphisation, increasing diffusion rates during subsequent calcination and promoting a faster perovskite-structure formation. This was followed by heat treatment in a tube furnace at 700°C for 2 hours at 3.7 rpm, to remove acetate ligands and form $\text{Ba}_2\text{Bi}_2\text{O}_6$. The mass of the sample was recorded before and after calcination in order to quantify the mass loss associated with ball milling and the removal of organic acetate species.

To evaluate the photothermal performance of $\text{Ba}_2\text{Bi}_2\text{O}_6$, the material was deposited on a filter paper and tested under simulated solar irradiation (under a light intensity of 1 sun). The surface temperature of the material directly reflected the photothermal conversion ability of the milled $\text{Ba}_2\text{Bi}_2\text{O}_6$, whilst water evaporation was the main indicator of the solar-driven interfacial evaporation performance. A total of 20mg was dispersed in 2ml of de-ionised (DI) water and subjected to ultrasonic treatment for 10 minutes to achieve a uniform

suspension. A 4cm x 8cm main filter paper was used as the supporting layer. A circular and uniform $\text{Ba}_2\text{Bi}_2\text{O}_6$ powder film with a 3.5cm diameter was prepared by deposition of the formed suspension on a filter paper and evaporating DI water on a hotplate at 75°C for 15 minutes.

In the following surface temperature and evaporation measurements, a Xenon lamp equipped with a AM1.5 solar filter was used as simulated solar light at 1sun (0.1 W cm^{-2}). The surface temperature measurements were carried out both on a dry and wet basis. For the wet-basis conditions, the filter paper was partially immersed in a 100 ml beaker with water to support the transportation of water along the filter paper to the evaporation interface. Water evaporation rate was measured by the change in mass of the bulk water.

A balance with high accuracy (A&D) was connected to a computer to record the mass change of water. To reduce the heat loss to the surroundings, a polystyrene box, as shown in *Figure 2.1.1A* was used to thermally insulate the $\text{Ba}_2\text{Bi}_2\text{O}_6$ powder film from the bulk water and the surrounding environment along with a quartz plate that was placed on the device above the material. A thermal couple was employed to record the temperature increase during both dry and wet measurements. The experiments were conducted for 30 minutes each, which was sufficient for the temperature to increase to the maximum and plateau. During the measurements, the room temperature was controlled at $22 \pm 1^\circ\text{C}$.

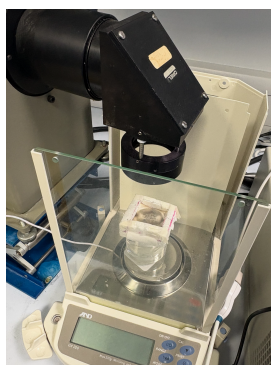


Figure 2.1.1A Setup of evaporation device.

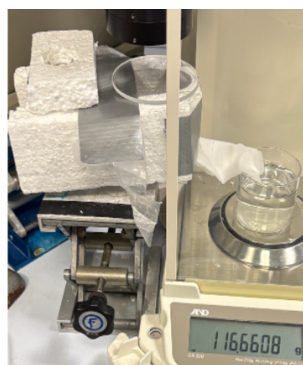


Figure 2.1.1B Setup of evaporation-condensation device.

2.2 Mass Loading Investigation

To determine the optimal mass load of this photothermal material for the given illuminated surface area of 9.62 cm^2 , a series of experiments were conducted, measuring the wet-basis surface temperature and the water evaporation rate for 30 minutes. Mass loadings of 20, 30, 40, 50 and 60 mg

of the $\text{Ba}_2\text{Bi}_2\text{O}_6$ powder were deposited on the filter paper using the method described in *Section 2.1*.

2.3 Condenser Design

A novel condenser device, aimed to promote efficient energy management, was fabricated from polystyrene and tested with the $\text{Ba}_2\text{Bi}_2\text{O}_6$ as illustrated in *Figure 1.2*. The device consisted of two compartments – an evaporation chamber and a condensation chamber, separated by a polystyrene wall to establish a temperature gradient and promote directional vapour transport and condensation. The condenser design was aimed to accelerate heat and mass transfer by maintaining a higher vapour pressure and temperature within the evaporation compartment and a lower temperature in the condensation compartment, thereby achieving an automatic and continuous evaporation–condensation cycle. The cold-water reservoir in the form of a beaker with cold water was placed in the condensation compartment establishes a stable temperature gradient, driving vapour transport from the hot to the cold side. It acts as a heat sink, absorbing the released latent heat and thereby accelerating droplet nucleation and condensation.

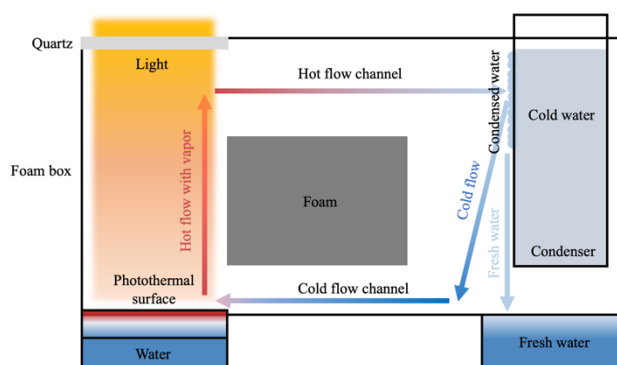


Figure 2.3 Diagram of evaporation-condensation device.

To confirm whether a thermodynamically stable framework was established within the condenser and to ensure an evaporation-condensation equilibrium is achieved to maximise the rate of mass transfer, temperatures profiles within the compartments were recorded. A total of 50mg of $\text{Ba}_2\text{Bi}_2\text{O}_6$ was deposited onto a 4cm x 14 cm filter paper, following the procedure outlined in *Section 2.1*.

The coated filter paper was placed into the device and one end was submerged in a beaker with water to provide a continuous water flow, as demonstrated in *Figure 2.4.1B*. Following the same procedure to measure the temperature and evaporation rate as stated in *Section 2.2*, the temperature at various points was recorded for 30 minutes each during device operation. This yielded

6 temperature readings: T0 - temperature on the Ba₂Bi₂O₆ surface (evaporation interface), T1 - temperature at the bottom of the evaporation compartment, T2 - temperature at the top of the evaporation compartment, T3 - temperature at the bottom of the condensation compartment, T4 - temperature at the top of the condensation compartment, T5 - Temperature of water in the beaker in the condensation compartment, used for accelerating droplet formation and enhancing condensation.

To further investigate the effectiveness of the system, extended experiments were conducted for 6-hour periods and left overnight. The wet basis temperature of the photothermal material and the evaporation rate were monitored throughout the duration of the experiments. In addition, the temperature of the water in the condensation compartment (T5) was recorded every 2 hours to track heat exchange and its influence on condensation. The condensed water was collected in a beaker at the end of each experiment.

2.4 Water Evaporation and Fresh Water Collection Stability

To compare the evaporation stability of DI water and saline water, a series of overnight experiments was carried out using the condenser setup described in *Section 2.3*. A seawater solution was prepared by dissolving 35.0 g of sodium chloride (NaCl) in 1.000 L DI water. Both DI water and the 35 ppt NaCl solution were used as feed solutions in separate runs. The mass of the bulk liquid was recorded continuously throughout the experiment, allowing a time-dependent evaporation rate and the total condensed mass to be recorded. The condensed water was removed from the device at the end of each experiment into a collection vessel and its volume was recorded. An energy balance was performed using the measured evaporated and condensed masses to quantify the fraction of the evaporation energy that was recovered during condensation to assess the effectiveness of the water reservoir within the condenser (heat sink) for both conditions.

2.5 Characterisation techniques

Powder XRD was used to determine the crystalline phases and assess phase purity of Ba₂Bi₂O₆ and the BaO/Bi₂O₃ reference samples. The results collected on a PANalytical Aeris diffractometer using Cu K α radiation ($\lambda = 1.54 \text{ \AA}$) and recorded over $2\theta = 20\text{--}70^\circ$. The diffraction peaks were then compared to the JCPDS reference patterns for Ba₂Bi₂O₆ (01-074-2124), BaO (26-0178) and Bi₂O₃ (76-1730) to evaluate the Ba₂Bi₂O₆ phase.

The optical properties of powders were evaluated using UV–Vis diffuse reflectance spectroscopy. The reflectance spectra were used to infer their relative light absorption ability and therefore assess the suitability of Ba₂Bi₂O₆ as a photothermal material for solar interfacial evaporation. The reflectance spectra (R, %) was recorded from 200 to 800 nm using Shimadzu UV–Vis spectrophotometer operated in reflectance mode. The powders were measured as packed beds.

3. Results and Discussion

3.1 Ba₂Bi₂O₆ Synthesis and Characterisation

Powder samples were prepared via a solid-state reaction were characterised by powder X-Ray diffraction (XRD) and UV-Vis spectroscopy to assess phase purity and confirm the formation of the Ba₂Bi₂O₆ double perovskite structure. As the precursor mixture undergoes heat treatment in a tube furnace in the presence of air, the formation of bismuth oxide and barium oxide (Bi₂O₃ and BaO respectively) is thermodynamically favourable and should be minimised.

To optimise synthesis, a series of Ba₂Bi₂O₆ batches were prepared with systematic calcination and ball-milling conditions. Initial batches were milled at 100 rpm for 10 minutes and calcined at 700 °C for 2, 3 and 4 hours. The resulting samples exhibited visible yellow and white impurities attributed to barium and bismuth oxides and their XRD pattern contained additional diffraction peaks assignable to these components. Increasing the calcination temperature to 950°C for 2 hours suppressed these impurity peaks, but resulted in over-sintering. The dry solid powder could not be uniformly deposited onto the filter paper for photothermal testing.

The ball-milling conditions were therefore optimised to promote the formation of a purer crystalline Ba₂Bi₂O₆ phase at lower calcination temperature. Whilst increasing the milling speed and duration enhances mixing and reactivity, it risks local overheating and premature reaction inside the mill. The most suitable powder was ultimately obtained at 500 rpm for 1.5h followed by calcination at 700°C for 2 hours. *Figure 3.1.1* demonstrates how the optimised Ba₂Bi₂O₆ was deposited on the filter paper.

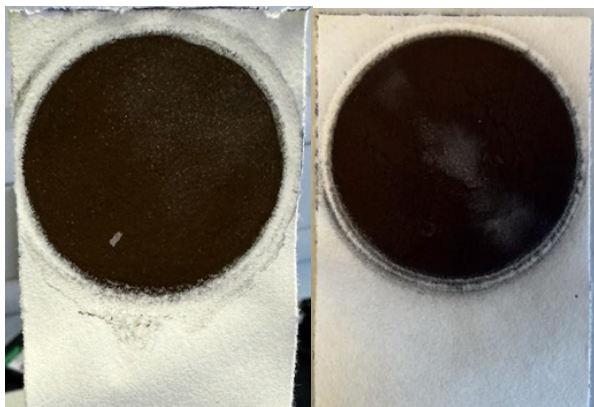


Figure 3.1.1 $\text{Ba}_2\text{Bi}_2\text{O}_6$ deposited on filter paper (higher mass loading on the left than of that on the right).

3.2 XRD Characterisation

Figure 3.2.1 shows the XRD pattern of the optimised $\text{Ba}_2\text{Bi}_2\text{O}_6$ sample (black) together with reference patterns for Bi_2O_3 (orange), BaO (blue) and $\text{Ba}_2\text{Bi}_2\text{O}_6$ (stick patterns, JCPDS 76-1730, 26-0178 and 01-074-2124). The prominent diffractions of the optimised sample can be indexed to the $\text{Ba}_2\text{Bi}_2\text{O}_6$ phase and no peaks attributable to Bi_2O_3 or BaO are detected within the resolution of the measurement. The main diffraction peaks at $2\theta = 21, 29, 42$ and 61 and 69 correspond to (110), (200), (211), (220) and (310). These coincide with the characteristic peaks for $\text{Ba}_2\text{Bi}_2\text{O}_6$ -based double perovskites and confirm that the optimised milling and calcination conditions yield a predominantly single-phase $\text{Ba}_2\text{Bi}_2\text{O}_6$ double perovskite suitable for subsequent photothermal and interfacial evaporation analysis.

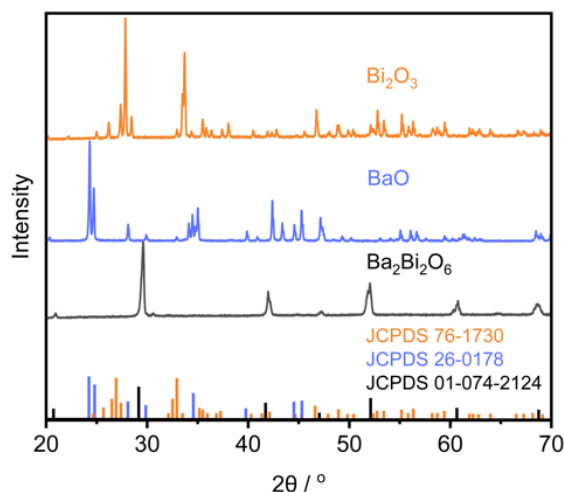


Figure 3.2.1 XRD of $\text{Ba}_2\text{Bi}_2\text{O}_6$, BaO and Bi_2O_3 .

3.3 UV-Vis Diffuse Reflection Characterisation

The optical properties of the synthesised $\text{Ba}_2\text{Bi}_2\text{O}_6$ samples were evaluated by the UV-Vis diffuse reflectance spectroscopy, and the optimised

$\text{Ba}_2\text{Bi}_2\text{O}_6$ material can be seen in Figure 3.3.1. The $\text{Ba}_2\text{Bi}_2\text{O}_6$ samples show the lowest reflectance across the entire 200-800 nm range and remain below 19%. On the contrary, BaO exhibits a moderate, higher reflectance that is gradually increasing, but remains below 28%. Bi_2O_3 displays a sharp rise from 420nm and reaches almost 100% reflectance at 600nm, after which it plateaus. The behaviour of the materials is consistent with the black appearance of $\text{Ba}_2\text{Bi}_2\text{O}_6$, the yellow colour of Bi_2O_3 and the grey colour of BaO , which is atypical for usual barium oxides that are white. Figure 3.3.2 demonstrates that $\text{Ba}_2\text{Bi}_2\text{O}_6$ is significantly darker than either of the binary oxides.

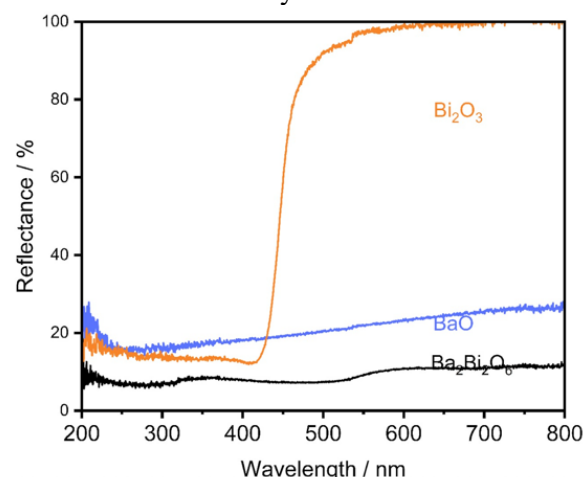


Figure 3.3.1 UV-Vis of $\text{Ba}_2\text{Bi}_2\text{O}_6$, BaO and Bi_2O_3 .

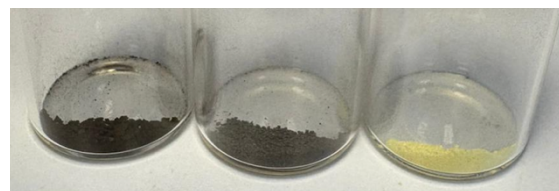


Figure 3.3.2 Appearances of $\text{Ba}_2\text{Bi}_2\text{O}_6$ (left), BaO (middle) and Bi_2O_3 (right).

For an optically thick, opaque powder, the transmitted light is negligible, therefore energy gives $A(\lambda) + R(\lambda) \approx 1$, where A is the spectral absorbance and R is the total reflectance^[4]. Therefore, a lower reflectance directly implies a higher absorption at that wavelength^[5].

The relationship between diffuse reflectance and absorption is commonly formulated using the Kubelka-Munk (K-M) remission function which is proportional to the absorption coefficient of powdered samples and can be used to analyse band gaps and absorption edges of semiconductor oxides^[6].

$$F(R) = \frac{(1-R)^2}{2R} = \frac{K}{S} \quad (\text{Eq. 1})$$

The $F(R)$ is the remission function, R is the reflectance ratio of the sample and the standard

material and K and S are the K-M absorption coefficients. Therefore, diffuse reflectance UV-Vis is particularly suitable for strong scattering, opaque powders, such as $\text{Ba}_2\text{Bi}_2\text{O}_6$, where conventional transmission and absorbance measurements are unreliable.

It can be concluded that the optimised $\text{Ba}_2\text{Bi}_2\text{O}_6$ phase absorbs significantly more light over the solar spectrum than either BaO or Bi_2O_3 . This is consistent with the ability of photothermal materials to exhibit strong absorption across the visible and near-infrared regions, which account for the majority of the incident solar power, to allow efficient solar-driven interfacial evaporation [7] [8].

3.4 Interfacial Temperature Change on a Dry Basis

The photothermal performance of $\text{Ba}_2\text{Bi}_2\text{O}_6$ was first evaluated by monitoring the surface temperature increase (ΔT) on a dry basis under 1 sun irradiation. Upon illumination, as seen in *Figure 3.4.1*, the optimised $\text{Ba}_2\text{Bi}_2\text{O}_6$ films show a rapid initial temperature rise, reaching $\Delta T \approx 40\text{--}45^\circ\text{C}$ within the first 5 minutes. This is followed by a quasi-steady value of $\Delta T \approx 48^\circ\text{C}$ after 20–30 minutes. The solid-red line represents the mean temperature increase obtained from independent measurements of optimal $\text{Ba}_2\text{Bi}_2\text{O}_6$ material and the shaded region corresponds to experimental variability. It can be concluded that the material effectively converts incident solar energy into heat energy and demonstrates a good stability of the photothermal response over the irradiation period, suggesting that the material will behave similarly for prolonged periods of time.

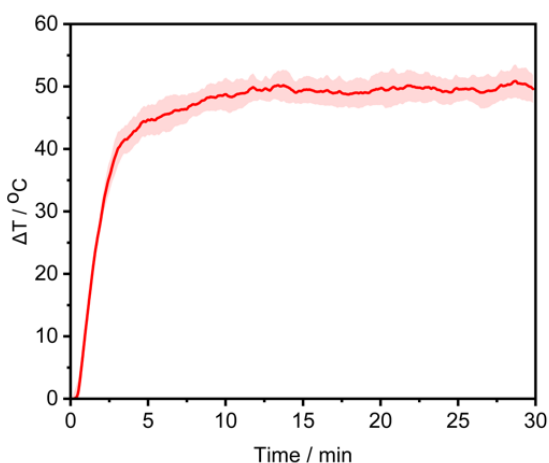


Figure 3.4.1 Average temperature change of mass loadings on a dry basis.

The magnitude of the temperature rise for this photothermal material is comparable to the existing photothermal absorbers. For example, under

similar conditions, carbon-based evaporators [9] typically exhibit temperature increase of $30\text{--}75^\circ\text{C}$ under 1 sun, depending on substrate and coating design. Carbon black/polymer composites reach surface temperatures of $60\text{--}70^\circ\text{C}$ [10]. Similarly, TiO_2 based photothermal systems demonstrate a more modest temperature increase. A TiO_2 /activated carbon (TiO_2/AC) evaporator could reach $\sim 35^\circ\text{C}$ within 6 minutes under 1 sun and overall reach approximately 40°C , whilst pure TiO_2 reaches 27°C in the first 6 minutes and 33°C overall [11]. It should also be noted that the greater initial temperature rise for $\text{Ba}_2\text{Bi}_2\text{O}_6$, suggest a faster photothermal response, which is consistent with a more efficient early-stage conversion of solar irradiation into heat energy. Furthermore, a black TiO_2 /PDMS sponge attains a steady surface temperature of $\sim 66^\circ\text{C}$, significantly higher than $\sim 42^\circ\text{C}$ achieved only by PDMS under the same conditions [12].

The literature values place the $T \approx 50^\circ\text{C}$ achieved by $\text{Ba}_2\text{Bi}_2\text{O}_6$ withing the performance parameters, supporting its use as an efficient, non-carbon oxide absorber for solar-driven interfacial evaporation. Additionally, oxide-based photothermal materials have the advantage of high chemical and thermal stability and resistance to photodegradation. Particularly the perovskite oxides show a tuneable electronic structure, which increases water stability and allows for band-gap adjustability through compositional design.

3.5 Optimisation of Mass Loading

When finding the optimal mass loading of $\text{Ba}_2\text{Bi}_2\text{O}_6$, the corresponding evaporation rates were calculated as shown in *Figure 3.5.1*. Increasing the photothermal mass on the filter paper from 20mg to 40mg results in a much steeper gradient and therefore a much higher evaporation rate of water, increasing by $0.28 \text{ kgm}^{-2}\text{min}^{-1}$. However, loadings greater than 40mg exhibited a much lower increase in evaporation with 50mg having a slightly higher rate than 60mg. The reason for this is that a plateau is reached at 40mg due to absorption saturation and adding excessive $\text{Ba}_2\text{Bi}_2\text{O}_6$ is no longer beneficial for the system as the unnecessary weight and thickness worsens heat transfer, taking longer to display the same mass change as lower the loadings.

Figure 3.5.2 illustrates the temperature change corresponding to the different masses. It can be concluded that 50mg exhibits the lowest surface temperature increase of approximately 17°C , the greatest heat energy uptake for the latent heat of

vaporisation of water among the other mass loadings. Consequently, it exhibits a higher energy efficiency. The second highest temperature increase is shown with 40mg whilst it can be implied that 60mg has much less thermal energy utilised for vaporisation, with $\Delta T \approx 21^\circ\text{C}$. This proves further that excessive $\text{Ba}_2\text{Bi}_2\text{O}_6$ loading causes added downsides that outweigh the benefits.

Figure 3.5.3 demonstrates the efficiencies of the five mass loadings investigated with the following equation used:

$$\eta = \frac{cm\Delta T + mH}{I\Delta t} \quad (\text{Eq. 2})$$

c = Specific heat of water ($4.184 \text{ Jg}^{-1} \text{ }^\circ\text{C}^{-1}$)

m = Evaporative mass change of water (g)

ΔT = Temperature Change ($^\circ\text{C}$)

H = Latent heat of vaporisation (2256 J/g)

I = Light intensity (0.1 Wm^{-2})

Δt = Time duration (seconds)

As expected, the efficiency associated with 20mg is much lower than that of the other results, whilst increasing the mass above 30mg shows a steady increase. However, it can be concluded that 50mg has the highest efficiency value at 0.96, which is greater than 60mg at 0.95. This is also a result of reduced heat transfer due to a thicker layer of $\text{Ba}_2\text{Bi}_2\text{O}_6$ and a reduced mass change of water. Although mass loadings higher than 60mg were not investigated, it can be predicted that both the evaporation rates and efficiencies will not increase further. Furthermore, it would lead to unnecessary increased material costs especially during scale-up of the experiment. Therefore, from these data, it was decided that the optimal mass to use for the device was 50mg.

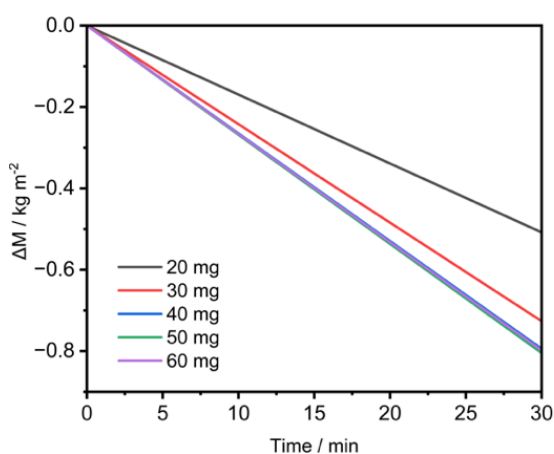


Figure 3.5.1 Evaporation rate of mass loadings.

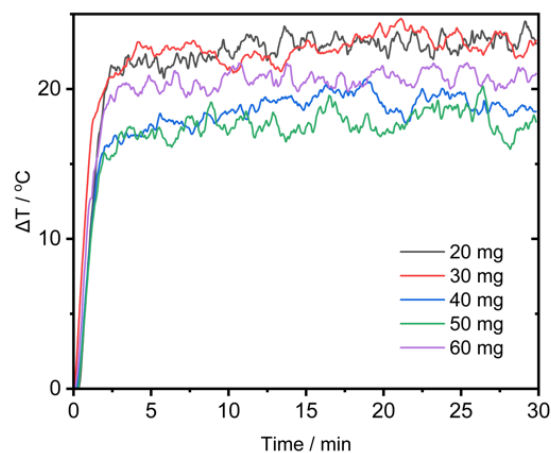


Figure 3.5.2 Temperature changes of mass loadings on a wet basis.

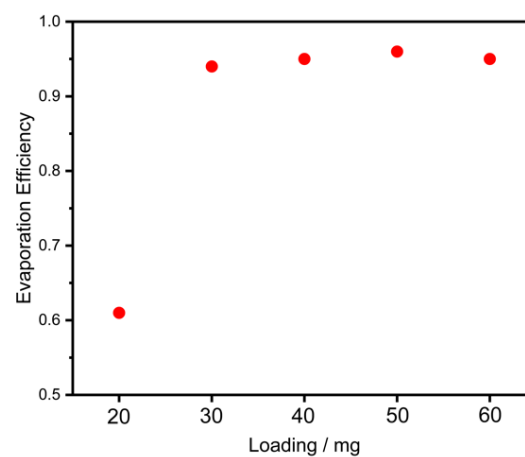


Figure 3.5.3 Efficiencies of mass loadings.

3.6 Device Evaporation Rate Using Condensation Device

Figure 3.6.1 illustrates the benefit of using the novel evaporation-condensation device. The evaporation rate in an open system (red line) is typically higher than in a glass-condenser closed system (blue line) due to saturated vapour being continuously removed from the evaporation surface. However, an open system risks the loss of the evaporated water and its latent heat to the surrounding. The objective of this closed system device is to operate like an open system and establish an efficient energy management, condensation and evaporation. Due to the less favourable conditions such as heat loss, increased humidity and vapour pressure, the device (black line) cannot achieve the exact evaporation rate as that of an open system. However, the device has been designed so that the water vapour is directed to the condensation compartment of a lower vapour pressure resulting in a net driving force where open system-like conditions are maintained in the evaporation compartment. Subsequently, this enables an evaporation rate close to that of an open system.

Additionally, temperature was measured at five different point within the device during operation, with the conclusion being that the evaporation compartment exhibited an internal temperature of over 30°C and in the condensation compartment a temperature of 24°C was recorded. The sustained temperature difference demonstrates that a vapour-pressure gradient is achieved and the device is operating successfully.

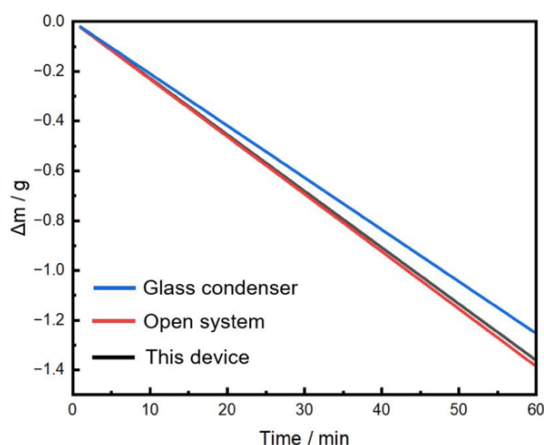


Figure 3.6.1 Evaporation rate compared to open and closed systems.

3.7 Water Evaporation and Fresh Water Collection Stability

The data from the overnight evaporation experiments for water and the 35 ppt NaCl solution are displayed in *Figure 3.7.1*. The evaporation rate of water (blue line) remained stable throughout, as shown by the linear decrease in mass change with time. Subsequently, 0.7g of water was collected against an evaporated mass of 44.1g, corresponding to a recovery of 1.59% of the total evaporated mass. Droplets were observed on the walls of the condensation device, the water reservoir in the condensation compartment and the quartz glass above the photothermal material, indicating that not all water was drained into the collection vessel. Therefore, further experiments are vital in treating and improving the condenser surface to allow more condensed water to be collected.

The red line shows the experiment repeated with seawater. As expected, its respective evaporation rate is not constant compared to water and gradually plateaus. This can be attributed to salt crystallisation formation on the evaporating surface, therefore reducing the effective surface area and increasing mass transport resistance. Additionally, the decrease in water (solvent) due to evaporation, leading to a higher salinity, results in a lower difference in vapour pressure, therefore diminishing the driving force. At the end of the

experiment, approximately 1g of freshwater was collected against an evaporated mass of 22.00g, corresponding to 4.55% recovery.

To further analyse the performance of the condenser, the following energy balances were performed for both experiments:

Energy collected in heat sink (J):

$$Q_{collected} = mc\Delta T \quad (\text{Eq. 3})$$

m = mass of cold water in beaker (g)

c = Specific heat of water (4.184 J g⁻¹ °C⁻¹)

ΔT = Temperature Change (°C)

Energy used for evaporation (J):

$$Q_{evaporated} = H\Delta m \quad (\text{Eq. 4})$$

Δm = Evaporative mass change of water (g)

H = Latent heat of vaporisation (2256 J/g)

Energy Recovered (%):

$$\eta = \frac{Q_{collected}}{Q_{evaporated}} \times 100 \quad (\text{Eq. 5})$$

This is used to determine the energies associated with evaporation and quantify the energy transferred to the cold-water reservoir in the condensation compartment during condensation. These were obtained using the latent heat of vaporisation and specific heat of water. The energy used for evaporation of water as the solvent is calculated to be 72,073.11J whilst for the saline solution it is 49,628.62J. The energy transferred to the cold-water reservoir was found to be 2111.75J for both cases, with its corresponding temperature increase being ~ 12°C. The energies recovered were found to be 2.93% for the water and 4.26% for the saline solution. Similarly to the mass analysis, these values highlight the low proportion of water that is returned in condensed form, emphasising the need for improvements of the cold water 'heat sink'.

Finally, energies provided by the 1-sun xenon lamp, used for latent heat of vaporisation and collected in the cold-water reservoir were recorded and plotted for both the water and saline water experiments, as displayed in *Figure 3.7.2*. Although the two experiments were run for marginally different times and the energy inputs were not the same, it can be shown that the latent heat of vaporisation of the saline water is lower than for that of water and so relatively less energy is required for its evaporation, highlighting its

greater energy recovery. Similarly, the lower latent heat of saline allows for slightly more efficient condensation on the surface of the heat sink, showing that this investigation is better suited to real life applications involving seawater than its lab-based counterpart involving DI water.

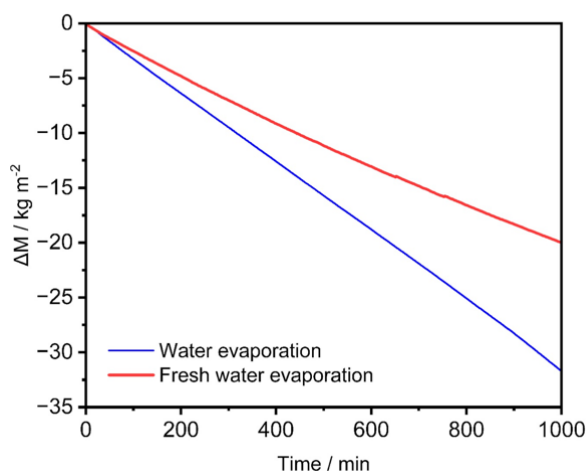


Figure 3.7.1 Evaporation rate of water and fresh water containing salt.

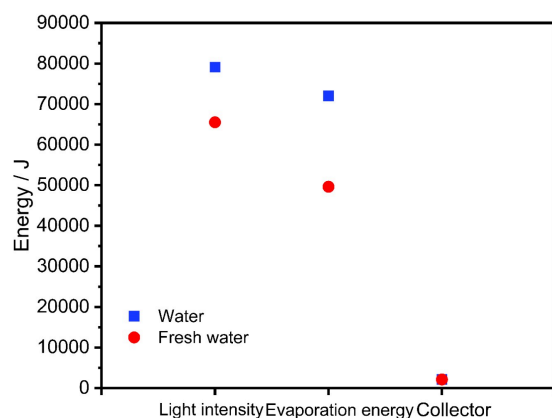


Figure 3.7.2 Light, evaporation and collected energies for water (blue) and saline water experiments (red).

4. Conclusion

This report set out to investigate the use of $\text{Ba}_2\text{Bi}_2\text{O}_6$ for clean water production through achieving its optimal synthesis conditions and loading mass for its use in the specially designed device containing evaporation and condensation compartments. Achieving these ideal conditions for synthesis consisted of high energy ball milling of barium acetate and bismuth acetate for 90 minutes at 500 rpm followed by calcination at 700°C for 2 hours in the tube furnace. The resulting $\text{Ba}_2\text{Bi}_2\text{O}_6$ product was then analysed and verified using XRD and UV-Vis which confirmed the formation of a high purity $\text{Ba}_2\text{Bi}_2\text{O}_6$ double perovskite phase and low reflectance across the solar spectrum indicated high absorption. This demonstrated its potential as an effective

photothermal material for solar interfacial evaporation.

Following this, the optimal loading mass was determined via measuring three main variables: temperature increase of $\text{Ba}_2\text{Bi}_2\text{O}_6$, the corresponding evaporation rates and efficiencies of samples between 20mg and 60 mg. It was concluded that 50mg was the preferred choice as efficiency began to plateau after 40 mg and excessive material on the filter paper added unnecessary thickness and weight, reducing heat transfer. The effectiveness of 50mg was investigated in the condenser device where it was utilised to condense and collect water from evaporating both DI water and saline solution. The mass of water recovered as a condensate was 1.59% and 4.55% respectively, as well as 4.54% and 6.60% was recovered in terms of energy.

The 96% efficiency of $\text{Ba}_2\text{Bi}_2\text{O}_6$ was compared to that of other photothermal materials obtained from previous research. It can be seen from Figure 4.1 that this material has a greater efficiency than titanium oxides and MXene, highlighting the effectiveness of $\text{Ba}_2\text{Bi}_2\text{O}_6$ as well as its associated benefits with regards to real-world applications and scale-ups.

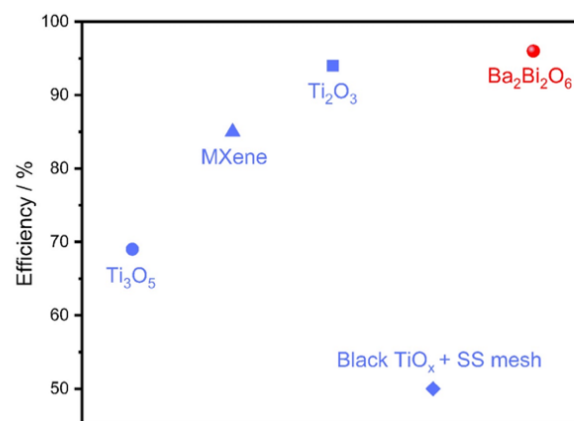


Figure 4.1 $\text{Ba}_2\text{Bi}_2\text{O}_6$ efficiency compared to other photothermal materials ^[13].

Additionally, although promising, future developments can be made to this research to ensure a greater rate of water collection. One way of achieving this is by modifying the device to improve the insulation and condensate collection methods. Further optimisation of the device to minimise heat loss and treatment of the condensation surface is required to maintain a larger temperature gradient. A better design focused on directing the condensed water to the designated outlet, rather than condensing on surfaces within the condenser would significantly

improve the performance. Further testing should also be carried out to determine whether a more optimal loading mass of $\text{Ba}_2\text{Bi}_2\text{O}_6$ is available. This can be done through several repeats, at smaller increments, between 40mg and 60mg. In addition, an analysis of the fresh water could be carried out to investigate whether the freshwater produced is not contaminated with $\text{Ba}_2\text{Bi}_2\text{O}_6$ as leaching is a potential risk when this system is scaled up for real world applications.

References

- [1] Yañez-Barnuevo, M. (2025). *Data Centers and Water Consumption* | Article | EESI. [online] Eesi.org. Available at: <https://www.eesi.org/articles/view/data-centers-and-water-consumption>.
- [2] Harvey, Fiona. “Global Fresh Water Demand Will Outstrip Supply by 40% by 2030, Say Experts.” *The Guardian*, 17 Mar. 2023, www.theguardian.com/environment/2023/mar/17/global-fresh-water-demand-outstrip-supply-by-2030.
- [3] [13] Cui, X., Ruan, Q., Zhuo, X., Xia, X., Hu, J., Fu, R., Li, Y., Wang, J. and Xu, H. (2023). Photothermal Nanomaterials: A Powerful Light-to-Heat Converter. *Chemical Reviews*, 123(11), pp.6891–6952. doi: <https://doi.org/10.1021/acs.chemrev.3c00159>.
- [4] Liu, M.-H., Hu, E.-T., Yao, Y., Zang, K.-Y., He, N., Li, J., Zheng, Y.-X., Wang, S.-Y., Yoshie, O., Lee, Y., Wang, C.-Z., Lynch, D.W. and Chen, L.-Y. (2014). High efficiency of photon-to-heat conversion with a 6-layered metal/dielectric film structure in the 250-1200 nm wavelength region. *Optics Express*, [online] 22(S7), p.A1843. doi: <https://doi.org/10.1364/oe.22.0a1843>.
- [5] Levinson, R., Chen, S., Ferrari, C., Berdahl, P. and Slack, J. (2017). Methods and instrumentation to measure the effective solar reflectance of fluorescent cool surfaces. *Energy and buildings*, 152, pp.752–765. doi: <https://doi.org/10.1016/j.enbuild.2016.11.007>.
- [6] Escobedo-Morales, A., Pedraza-Chan, M.S., Ruiz-López, I.I., Anot, E.C., Salazar-Villanueva, M. and Cortés-Arriagada, D. (2025). Accurate determination of the band gap energy of non-translucent semiconductor materials through the Tauc method: Theoretical framework, limitations, technical hints, and automated algorithms. *Next Materials*, [online] 10, p.101412. doi: <https://doi.org/10.1016/j.nxmate.2025.101412>.
- [7] Ji, S., Zhang, Z., Zhang, M., Yang, Z., Fan, Y., Zhang, J., Pang, Y. and Cui, L. (2025). Interfacial Solar Evaporation for Treating High-Salinity Wastewater: Chance and Necessity. *Processes*, [online] 13(9), pp.2679–2679. doi: <https://doi.org/10.3390/pr13092679>.
- [8] Xiao, Y., Guo, H., Li, M., He, J., Xu, X., Liu, S., Wang, L. and James, T.D. (2025). Strategies for enhancing the photothermal conversion efficiency of solar-driven interfacial evaporation. *Coordination Chemistry Reviews*, 527, p.216378. doi: <https://doi.org/10.1016/j.ccr.2024.216378>.
- [9] Guo, Y., Li, C., Wei, P., Hou, K. and Zhu, M. (2022). Scalable carbon black deposited fabric/hydrogel composites for affordable solar-driven water purification. *Journal of Materials Science & Technology*, 106, pp.10–18. doi: <https://doi.org/10.1016/j.jmst.2021.07.032>.
- [10] Sun, S., Sun, B., Wang, Y., Fordjour Antwi-Afari, M., Mi, H.-Y., Guo, Z., Liu, C. and Shen, C. (2021). Carbon black and polydopamine modified non-woven fabric enabling efficient solar steam generation towards seawater desalination and wastewater purification. *Separation and Purification Technology*, [online] 278, p.119621. doi: <https://doi.org/10.1016/j.seppur.2021.119621>.
- [11] Wei, Z., Muhammad Sultan Irshad, Arshad, N., Noureen, L., Ahmed, I., Mushtaq, N., Muhammad Sohail Asghar, Hayat, Q., Uzma Ghazanfar, Idrees, M., Shahzad, N. and Lu, Y. (2022). Nanocomposite-Enhanced Efficient Evaporation System for Solar-Driven Seawater Desalination—An Optimized Design for Clean Water Production. *Nanomaterials (Basel)*, 12(19), pp.3296–3296. doi: <https://doi.org/10.3390/nano12193296>.
- [12] Keerthnasre Dhandapani, Hossein Fattahimoghaddam, Kim, I.H., Mani, D., Kim, S.R., An, T.K. and Jeong, Y.J. (2024). Black TiO₂-Infused Polydimethylsiloxane Composites for Efficient Solar-Assisted Water Evaporation and Thermoelectric Energy Generation. doi: <https://doi.org/10.2139/ssrn.4704969>.

Temperature Dependence of Early-Age CO₂ Sequestration in Coke-Modified Concrete

Keziah Amuwo and Minsun Oh

Department of Chemical Engineering, Imperial College London, U.K.

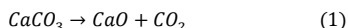
Abstract

Unpublished work carried out by Dimitrios Nikolis, under the supervision of Professor Klaus Hellgardt, investigated the effect of adding modified coke to concrete at constant room temperature and found that a 10% sand replacement achieved the highest CO₂ uptake. Building on this foundation, this project examines how temperature influences the initial rate of carbonation in a coke-modified concrete mixture. A reference sample, with no modified coke, was also analysed to confirm the correlation between the addition of modified coke and increased uptake. Initial carbonation rates at 24, 27, and 30°C over a 1-day curing period were 8.767 ppm·min⁻¹, 4.370 ppm·min⁻¹, and 4.162 ppm·min⁻¹, respectively, indicating a negative relationship between curing temperature and initial carbonation rate. This relationship differed from the hypothesised positive relationship between temperature and carbonation shown in our MATLAB carbonation model.

1. Introduction

The release of carbon dioxide emissions is the largest human-caused contributor to global warming, with the concentration of atmospheric CO₂ increasing by 51% greater than its pre-industrial level [1]. Figure 1 displays the top four sectors that contribute to global CO₂ emissions, and of the 13% that stem from the manufacturing and construction sector [2], 8% of global CO₂ emissions are released by the concrete industry [3], primarily due to the cement-making process.

CO₂ emissions from cement production primarily contribute to Scope 1 emissions (direct emissions from combustion and reactions, such as calcination) and Scope 2 emissions (indirect emissions resulting from the purchase of electricity). Calcination is the main reaction responsible for CO₂ production in the cement industry as it is used in the production of clinker – a component of cement. Through the calcination, limestone (CaCO₃) is converted to calcium oxide under 900°C heat, and CO₂ is produced as a by-product.



China is the largest cement producer, producing 2.1 billion tonnes in 2023 [4], and emitting 718 million metric tonnes of CO₂ from cement production in the same year. Putting this project in a local context, the UK produced 7.7 million tonnes of cement in 2023 [5] and emitted 6.6 million tonnes of CO₂ in the same year [6]. The cement industry emits approximately 0.9kg of CO₂ per kg of cement, with most emissions stemming from the calcination and combustion of fuels used to heat kilns [7]. With global cement production estimated at approximately 4 billion tonnes annually [8], the concrete industry is a significant contributor to human-made CO₂ emissions. However, through a process called carbonation, concrete can uptake CO₂ during and after curing.

Carbonation is a natural electrochemical reaction that occurs during and after curing, in which carbon dioxide dissolves in the pore moisture of concrete to form carbonic acid. This carbonic acid then reacts with calcium hydroxide in concrete to form calcium carbonate. Carbonation is generally a slow process in unmodified concrete mixtures. Due to the substantial amount of CO₂ that cement production emits, it is crucial that carbonation be maximised to offset the negative environmental impact of cement production.

Researchers have made significant strides in investigating sustainable concrete pathways, primarily through the improvement of cement plants, CCS techniques, and the use of modified coke to increase total carbonation. Although coke has been shown to influence carbonation at room temperature, no studies have evaluated its temperature sensitivity during early-age curing. This study aims to achieve this by utilising various temperature scenarios to provide regional recommendations on which countries should implement the modified coke addition in their concrete. This study also aims to confirm the hypothesis that coke-modified concrete increases carbonation due to increased porosity.

2. Background

2.1. Mechanisms for CO₂ absorption by concrete

Before CO₂ from the air enters the concrete, the CO₂ gas in the bulk air must first move through the air via diffusion and convection to reach the concrete surface. For simplicity, it is assumed that gases in the bulk phase are well-mixed. This assumption is justified using a method employed by Arnold et al., who confirmed the “well-mixed room” model by measuring contaminant concentrations at six different chamber locations [9]. A similar approach was employed in this study, where the CO₂ sensor in our experiment was used to measure CO₂ concentrations at various locations within the experimental glove box. The results confirmed that

the CO₂ concentrations were approximately the same across different locations. This assumption implies that we can use the Dirichlet boundary condition in which the CO₂ concentration at the concrete surface is equal to the bulk concentration, allowing the focus of this study to be on the transport of CO₂ through the concrete pores.

Once the CO₂ reaches the concrete surface, the gas will penetrate the concrete through air-filled pores. This diffusion process is driven by the concentration gradient between the bulk CO₂ and the pore CO₂ concentration, and is governed by Fick's second law:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \quad (2)$$

which models the concentration change of CO₂ over time due to diffusion. Li et al. developed various mathematical models for CO₂ transport through the concrete pore network, depending on the relative speeds of carbonation and CO₂ diffusion [10]. The MATLAB model will utilise equations from Li et al.'s work, together with equations derived by AL-Ameeri et al. that describe the effect of climate change on carbonation [11].

The model will utilise the effective diffusivity expression from AL Ameeri et al.'s work, which is written as the product of three correction factors that account for the influences of temperature (T), relative humidity (RH), and concrete composition. Changes in concrete composition, as influenced by mix parameters such as the water-to-cement ratio and replacements, affect the pore structure and crack widths (W_c).

$$D_{CO_2} = D_{CO_2,ref} f_{C1}(T) f_{C2}(RH) f_{C3}(W_c) \quad (3)$$

$$f_{C1}(T) = e^{\frac{U_c}{R} \left(\frac{1}{T_{ref}} - \frac{1}{T} \right)} \quad (4)$$

$$f_{C2}(RH) = (1 - RH)^{2.2} \quad (5)$$

$$f_{C3}(W_c) = 11.4 \sqrt{W_c} + 1 \quad (6)$$

AL-Ameeri et al.'s effective diffusivity will then be introduced into Li et al.'s expression for a diffusion front ($\xi(t)$) that moves with time.

$$\xi(t) = k_\xi \sqrt{D_{CO_2} t} \quad (7)$$

$$k_\xi = \sqrt{\frac{2C_s}{S_{max}}} \quad (8)$$

k_ξ is a dimensionless constant that can be obtained by applying a mass balance at the moving diffusion front. Li et al. approximated this dimensionless constant by solving the mass balance, as seen in Equation 8. C_s is the surface CO₂ concentration, and

S_{max} is the maximum content of CO₂ that can be consumed in carbonation (kg/m³).

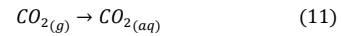
To ensure the carbonation depth does not exceed the thickness of the concrete sample, the following expression can be used to represent the effective carbonation depth:

$$x_c(t) = \min[\xi(t), h] \quad (9)$$

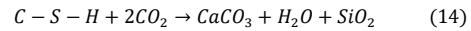
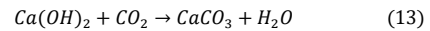
Where h is the thickness of the concrete. The total CO₂ uptake ($M_{CO_2}(t)$) is found by multiplying the carbonated volume by the maximum content of CO₂ that can be consumed by carbonation:

$$M_{CO_2}(t) = AS_{max} x_c(t) = AS_{max} \min \left[k_\xi \sqrt{D_{CO_2} t}, h \right] \quad (10)$$

Once the CO₂ gas has diffused through the concrete pore network, it undergoes solvation and becomes aqueous CO₂ in the liquid phase of the concrete. The aqueous CO₂ undergoes hydration to become carbonic acid, and carbonic acid undergoes ionisation to form carbonic species.



These aqueous equilibria will maintain carbonate species in the pore solution. The H⁺ ions cause a drop in pH of the cementitious system, which is why steel reinforcement, used in industry with concrete, corrodes over time [12]. Dissolution of portlandite ($Ca(OH)_2$) and C-S-H occurs rapidly, releasing ions (primarily Ca^{2+}) into the solution near the carbonation front. Carbonate ions react with calcium cations to precipitate calcium carbonate which reduces porosity and slows further diffusion of carbon dioxide into the concrete. The primary carbonation reaction is illustrated in Equation (9).



C-S-H is also prone to decalcification when carbon dioxide is present, forming calcium carbonate and silica gel (SiO_2) [12].

2.2. Concrete matrix composition

Several studies have been conducted to find the optimal concrete composition that provides adequate compressive strength for construction. The main materials in concrete are a binder (such as cement), aggregate (such as sand), and water. Cheah et al. conducted experiments to optimise the ratios of sand, cement and water in concrete in terms of overall workability and compressive strength. The study concluded that a cement-to-sand ratio of 1:2.25, combined with a water-to-cement ratio within a range of 0.40–0.50, yielded maximum

workability and compressive strength [13]. As this study focuses on maximising the rate of carbonation, it is important to understand how the matrix composition affects carbonation. Mehdizadeh et al. discussed the effect of the water-to-cement ratio on the carbonation rate, concluding that mixes with higher ratios yielded greater carbonation due to increased hydration and higher porosity [14]. Increased porosity allows greater diffusion of CO₂ into the concrete. Marangu et al. reviewed the effect of internal and external factors on carbonation resistance, citing that Ordinary Portland Cement (OPC) is more resistant to carbonation than blended cements, and is thus less preferred [15]. The most significant difference between OPC and blended cement is that OPC consists mainly of clinker, whereas blended cement is produced by replacing a portion of clinker with supplementary cementitious materials (SCMs) such as fly ash or slag. Clinker is a component of cement with high concentrations of Ca(OH)₂, which is a primary hydration product. The Ca(OH)₂ acts as a buffer in concrete, neutralising H⁺ ions formed from hydrated carbonic acid. However, in blended cements, the replacement of some of the clinker with SCMs reduces calcium hydroxide content, allowing the carbonation front to progress deeper into the concrete structure.

2.3. The effect of modified coke on concrete mixes

Pyrolytic carbon “coke” is a carbon material produced through the decomposition of hydrocarbon gases at high temperatures in an anaerobic environment. The addition of raw fine coke particles, as a replacement of aggregate, to the concrete matrix in theory would cause a net increase in the carbonation rate. This is because it behaves as an inert, porous carbon phase with a relatively high surface area, thereby increasing the overall porosity of the concrete mix, allowing CO₂ to more easily diffuse into the concrete pores. Parallels can be drawn between the addition of raw coke to the concrete mix and the addition of fly ash to the concrete mix. Jaworska et al. carried out an investigation into the effect of the addition of fly-ash to concrete on carbonation, and their results exhibited an increase in carbonation rate of concrete with the addition of fly ash, attributing this increased carbonation rate to the lower amount of Ca(OH)₂ and increased porosity that fly-ash provided [16].

Unpublished work done by Professor Klaus Hellgardt and Dimitrios Nikolis at Imperial College London investigated the optimal replacement of sand or cement in concrete with modified coke to optimise carbonation rate. After testing different replacement percentages of sand and cement, it was found that 10% replacement of sand by heat-treated modified coke yielded the greatest carbonation. It is this study that we plan to build upon.

Thermal treatment of coke results in changes to its physical structure, surface chemistry, and reactivity. Whilst undergoing heat treatment, the coke undergoes oxidation in which surface hydrophilicity and surface area of the coke are increased. The increased hydrophilicity of the coke, when incorporated into the concrete mix, increases the mix’s ability to hold water, which in turn can increase porosity and make CO₂ transport easier. Amorphous carbon, such as coke, will begin oxidation in air at a temperature greater than 300°C. For a coke surface to carry several oxygen-containing functional groups (e.g. carboxyl and phenolic groups), oxidation must occur at temperatures below 400°C. Upon oxidation, a number of oxygen-containing functional groups are attached to the carbon material structure via covalent bonding [17].

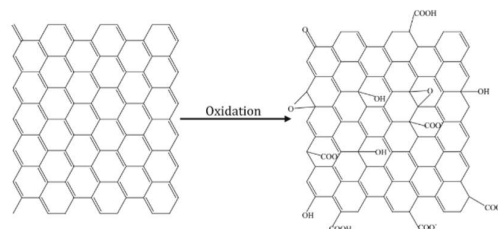


Figure 3. Diagram displaying the change in chemical structure of carbon material after oxidation. [17]

D.B. Fischbach conducted a study on the effect of high-temperature heat treatment on the structure of pyrolytic carbons, observing that the microstructure of the carbons faced very little change. However, he did observe that heating the carbon to high temperatures transformed its structure from a highly disordered, turbostratic structure to more graphitic, better-stacked crystallites with a more ordered microstructure. This, in turn, lowers surface reactivity and reduces the number of oxygen-containing functional groups, making the carbon particles less reactive and hydrophilic. Whilst the temperature range for this study’s heat treatment varies significantly from that in our study, Fischbach’s results highlight the importance of not using excessively high temperatures for coke treatment in the interest of increasing carbonation. [18]

2.4 External factors affecting carbonation

External factors, such as atmospheric CO₂ concentration, relative humidity, temperature, and curing period, influence the rate at which CO₂ penetrates the concrete matrix. Internal factors, such as material composition ratios, material type and content, and concrete porosity, tend to affect the resistance of the concrete to carbonation. As internal factors, such as material composition ratios, material type, and porosity, have already been discussed in this report, this section will focus on understanding

the effects of external factors on the carbonation of concrete.

2.4.1. Relative humidity

Relative humidity has a significant impact on carbonation, primarily due to its effect on pore hydration. A review by Alshaeer et al. reported that the carbonation process is most active when the RH is between 50-70% [12]. Pu et al. conducted a similar study on accelerated carbonation technology for the treatment of recycled concrete aggregates, confirming that this RH range was optimal for carbonation [19]. At low humidity levels, the carbonation reaction is restricted by a lack of moisture and pore hydration. In contrast, high humidity levels cause the pore network to become oversaturated with water, blocking the diffusion of CO₂ gas into the concrete.

2.4.2. Temperature

Changes in temperature can affect solubility, reactant transport speed, vapour condition, and hydration in concrete structures, all of which affect concrete carbonation [20]. The carbonation reaction kinetics are temperature sensitive.

Most investigations into the effect of temperature have been reflected in a change in carbonation depth. For example, Yoon et al. [21] confirmed that the carbonation rate and temperature are consistent with the Arrhenius law (seen in Equation 15).

$$k = Ae^{-\frac{E_a}{RT}} \quad (15)$$

Carbonation depth is often written as

$$x = k\sqrt{t} \quad (16)$$

Therefore, by combining Equation 15 and Equation 16 for two different temperatures, you obtain:

$$\frac{x_1}{x_2} = e^{U\left(\frac{1}{T_2} - \frac{1}{T_1}\right)} \quad (17)$$

where U is the ratio of E_a and R, and x₁ and x₂ are the carbonation depths at temperatures T₁ and T₂ respectively.[20]

2.4.3. Curing period

Despite longer curing periods increasing the strength of concrete, it decreases the carbonation depth because it allows the concrete to further hydrate, densifying the microstructure of the concrete and reducing porosity, which in turn reduces the permeability of the concrete [22]. Decreased permeability of concrete restricts CO₂ diffusion into the concrete structure, thus limiting carbonation.

2.4.4. CO₂ concentration

Generally, higher CO₂ concentrations increase both the carbonation rate and the carbonation depth due to a greater concentration gradient, which drives faster diffusion of CO₂ into concrete. However, very high CO₂ levels can lead to a change in the concrete's microstructure, making the concrete's surface layer denser and reducing the rate of CO₂ diffusion into the concrete over time. An experimental study conducted by Cui et al. confirmed the trend in carbonation depth increasing with CO₂ concentration [23]. The study also confirmed the levelling off of carbonation rate and depth as the CO₂ concentration reaches high levels.

3. Materials and methodology

This study employed an experimental approach to investigate the effect of temperature on CO₂ absorption during the early stage of curing. The aim was to determine how elevated curing temperature affects the rate and extent of CO₂ uptake under controlled atmospheric conditions. All experiments were conducted in a sealed glove box to measure changes in CO₂ concentration over a 24-hour period. This duration was chosen to capture the initial, high-rate phase of carbonation kinetics in the fresh concrete samples. Three curing temperatures were tested: 24°C, 27°C and 30°C, representing normal and moderately elevated curing environments. Two replicate experiments were performed for each condition.

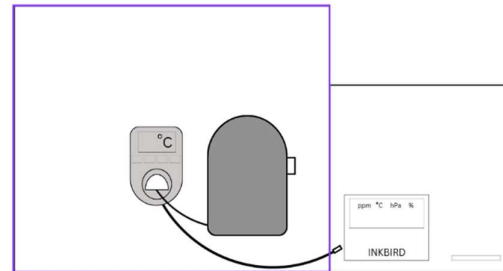


Figure 5. The thermocouple (far-left) is equipped with a temperature sensor that has been fed into the small glove box. The heater is plugged into the thermocouple and turns on when the pre-set 'on temperature' is reached. When the pre-set 'off temperature' is reached, the heater automatically switches off. Inside the small glove box, there is the CO₂ sensor and the concrete sample.

3.1 Materials

The raw materials used were sand, Type I Ordinary Portland Cement (OPC), deionised water, and coke. Sand acted as the fine aggregate, cement as the binder, and deionised water as the mixing fluid. Coke was introduced as a sand replacement to evaluate its influence on the carbonation behaviour of the concrete. CO₂ concentration and humidity changes were monitored using a Non-Dispersive Infrared (NDIR) CO₂ sensor placed inside a sealed glove box, along with a heater and thermocouple for temperature control. A schematic of the glove box

setup, including the CO₂ sensor and temperature control arrangement, is shown in Figure 5.

3.2 Experimental setup

Sand was dried in the HP 5890 Series II oven at 100 °C for 24 hours to eliminate moisture and prevent interaction within the concrete during curing. For the sand replacement mixture, sand was sieved to isolate particles $\leq 250 \mu\text{m}$, and 9 g of fine sand was removed and replaced with an equal mass of coke. The remaining sand fractions were recombined to ensure a uniform grain distribution. Coke was milled using the Fritsch Mortar Grinder Pulverisette 2 and then sieved using a 250 μm sieve to maintain the size distribution. The coke was thermally treated at 350 °C for 3 hours in the oven. In the oven, adsorbed moisture and volatile organic compounds were also removed.

Control and sand replacement samples were prepared. The composition of each mixture is shown in Table 1. All components were weighed using a balance and mixed in a beaker using the hand mixer to achieve uniform dispersion. The mixtures were then poured into square silicone moulds, levelled to eliminate air pockets. The mixture was covered with wet paper to maintain a humid curing environment. The thin geometry of the silicone moulds (87 mm $\times$ 87 mm $\times$ 9 mm) was used to ensure effective gas penetration through the concrete layer, uniform exposure, and minimised diffusion limitations during the experiment.

Table 1. Compositions of concrete.

Sample Type	Sand (g)	Cement (g)	Coke (g)	Water (g)
Control	90	40	0	20
Coke Added	81	40	9	20

After moulding, the samples were placed inside a sealed glove box equipped with a CO₂ sensor, heater, and thermocouple. The glove box atmosphere initially contained approximately 530 ppm CO₂, representing ambient air. To study the behaviour of coke-modified concrete at different temperatures, curing was conducted at 24 °C, 27 °C and 30 °C to evaluate the temperature dependence of early-age carbonation. The temperatures were selected to simulate various curing environments, depending on the specific curing location.

Additionally, to investigate the influence of carbon addition, two types of mixtures were prepared: a control concrete and a concrete modified with coke. Both mixtures were subjected to identical curing and monitoring conditions, allowing for a direct comparison of CO₂ uptake behaviour with and without carbon modification.

Humidity within the glove box was aimed to be maintained at 80%. Relative humidity was continuously monitored using an INKBIRD IAM-T1 to ensure compliance with the target humidity level. CO₂ monitoring started immediately after placing the moulds inside the glove box, providing continuous concentration data throughout 24 hours.

3.3. Leak test and correction

To quantify leakage from the external environment due to imperfect sealing, a leakage test was carried out. After each experiment, the CO₂ concentration inside the glove box was lower than the ambient atmospheric level. The concrete sample was transferred to an adjacent glove box through the connected port to prevent the introduction of outside air. The original glove box was then resealed, and the CO₂ concentration inside the empty glove box was monitored to determine the leak rate.

The rate of leakage was calculated by finding the straight line of best fit. After calculating the leakage rate, absorption measurements were corrected by subtracting the leakage contribution. The corrected concentration at any time t is:

$$C_{corrected}(t) = C_{measured}(t) - (r_{leak} \times t) \quad (18)$$

where t is time in minutes, r_{leak} is the leakage rate in ppm/min, $C_{corrected}$ is the corrected CO₂ concentration and $C_{measured}$ is the measured CO₂ concentration.

3.4 CO₂ sensor calibration

The NDIR CO₂ sensor was calibrated with fresh air before each experiment by placing it in well-ventilated outdoor air and confirming that it measured the expected ambient CO₂ concentration of approximately 430 ppm.

3.5 Data analysis

The recorded CO₂ concentration data were exported from the sensor software in .csv format and processed using MATLAB R2025b. Raw signals were first smoothed using a moving-average filter with a fitting window size of 5 and subsequently refined with a Savitzky–Golay filter with a fitting window size of 9 to reduce high-frequency sensor noise. For each dataset, the initial rate was obtained by performing a simple linear regression of the form:

$$C(t) = r_{initial} \times t + C(0), \quad t \leq 20 \quad (19)$$

where $r_{initial}$ represents the slope of the best-fit line. The slope is interpreted as the initial rate of concentration change (in ppm·min⁻¹). A more negative slope corresponds to faster initial CO₂ uptake. A 20-minute window was chosen to evaluate the initial absorption rate, taking into account short-term fluctuations at the start of the experiment.

To quantify the CO₂ concentration when the concentration profile reached a steady state, the time taken to achieve the final concentration was first calculated. Due to the CO₂ signal exhibits fluctuations, the time derivative of the smoothed concentration curve was used as an indicator of when the rate of change approached zero. A threshold of 0.01 ppm·min⁻¹ was selected as the criterion for steady-state behaviour. This value corresponds to the point where changes in concentration become smaller than the sensor's noise level and therefore no longer represent meaningful CO₂ uptake. The time taken until final concentration was then defined as the first time point at which a change in concentration is less than 0.01 ppm·min⁻¹.

Once the final concentration region was identified, the final CO₂ concentration was calculated as the mean value of the concentration over all time for this region, and the variability in this region was also quantified using the standard deviation. Equations 20a and 20b were used to calculate the concentration and its standard deviation. C_{final} is the final concentration, $C(t)$ is the concentration at the time t , and t_{final} is the time taken until the final concentration is achieved. Finally, total uptake of CO₂ is defined as the difference between initial and final concentrations.

$$C_{final} = \text{mean}[C(t) | t \geq t_{final}] \quad (20a)$$

$$\sigma_{final} = \text{std}[C(t) | t \geq t_{final}] \quad (20b)$$

A linear regression was applied again to calculate the rate of absorption, r_{final} after t_{final} (Equation 21).

$$C(t) = r_{final} \times t + C(t_{final}), t \geq t_{final} \quad (21)$$

Since the CO₂ concentration changes very slowly in this region, r_{final} is typically close to zero. Small deviations arise from sensor noise, micro-leakage, and small variations in equilibrium behaviour. The sensor accuracy was specified by the manufacturer (± 30 ppm + 3% of reading), while replicate variability of duplicate was quantified as the standard deviation between the two replicate runs. The uncertainties were combined using the root-sum-of-squares method, where U_{total} is total uncertainty, U_{sensor} is uncertainty of sensor, U_{rep} is uncertainty due to replicate variability.

$$U_{total} = \sqrt{U_{sensor}^2 + U_{rep}^2} \quad (22)$$

4. Results

This section presents the early-age CO₂ absorption behaviour of the concrete samples. First, the typical CO₂ concentration profiles and key metrics common to all experiments are summarised. Then, the effects

of coke addition at 24 °C and of curing temperature (24, 27 and 30 °C) on initial uptake rate, total uptake and time to steady state are described. Finally, the kinetic model fits and associated uncertainties are reported.

4.1 Overall CO₂ absorption behaviour

Across all experiments, the initial CO₂ concentration inside the glove box was approximately 530 ppm, with small run-to-run variation. The concentration decreased rapidly before t_{final} , followed by a slower approach to a nearly constant value over the remaining curing period.

Table 2 summarises the average and standard deviation of absorption rates, concentrations, total uptake and t_{final} over all runs. The highest absolute absorption rates ($r_{initial}$) occurred within the first 20 min, whereas beyond t_{final} the rate approached zero and the concentration remained within experimental uncertainty of the final value.

Table 2. Average CO₂ absorption behaviour of all experiments.

All experiments	Average	Standard deviation
$r_{initial}$ (ppm·min ⁻¹)	-2.965	0.628
r_{final} (ppm·min ⁻¹)	-0.042	0.017
$C(t = 0)$ (ppm)	533.750	10.340
C_{final} (ppm)	15.964	2.606
Total uptake (ppm)	517.786	12.528
t_{final} (min)	265.250	38.785

4.2 Effect of coke addition at 24 °C

At 24 °C, the coke-modified concrete exhibited a more rapid initial decrease in CO₂ concentration than the control. The initial absorption rate increased in magnitude from -4.071 ppm·min⁻¹ for the control to -8.767 ppm·min⁻¹ for the coke-modified sample, while the total CO₂ uptake over 24 hours remained similar for both mixtures.

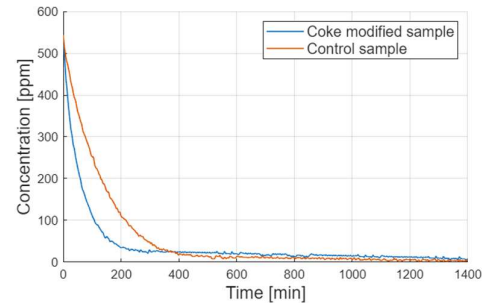


Figure 6. CO₂ concentration decay of control and coke-modified concrete at 24°C.

The time required to reach the plateau region was shorter for the coke-modified concrete than for the control. Figure 5 shows the CO₂ concentration profiles for both samples.

Table 3. Effect of coke addition on CO₂ absorption at 24 °C.

Sample type	Control	Modified
$r_{initial}$ (ppm·min ⁻¹)	-4.071	-8.767
r_{final} (ppm·min ⁻¹)	-0.026	-0.016
$C(t = 0)$ (ppm)	540.000	530.000
C_{final} (ppm)	12.569	17.434
Total uptake (ppm)	527.431	512.566
t_{final} (min)	255	225

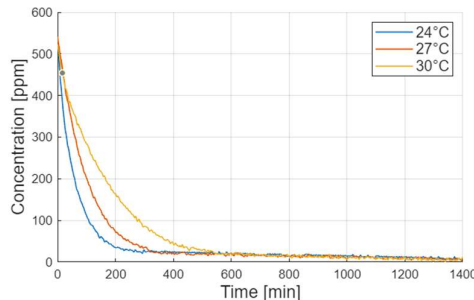
4.3 Effect of temperature on carbonation rate

The influence of curing temperature on CO₂ absorption was assessed at 24, 27 and 30 °C for coke-modified concrete. Increasing the curing temperature from 24 to 30 °C led to a decrease in the initial absorption rate and an increase in the time required to reach the final concentration as shown in Table 4.

Table 4. Absorption rate, CO₂ concentrations and t_{final} at different temperatures.

Temperature	24°C	27°C	30°C
$r_{initial}$ (ppm·min ⁻¹)	-8.767	-4.370	-4.162
r_{final} (ppm·min ⁻¹)	-0.016	-0.018	-0.039
$C_{initial}$ (ppm)	530.000	544.000	521.000
C_{final} (ppm)	17.434±5.6 56	15.373±6.9 03	18.481±14. 940
Total uptake (ppm)	512.566	528.627	502.519
t_{final} (min)	225	263	318

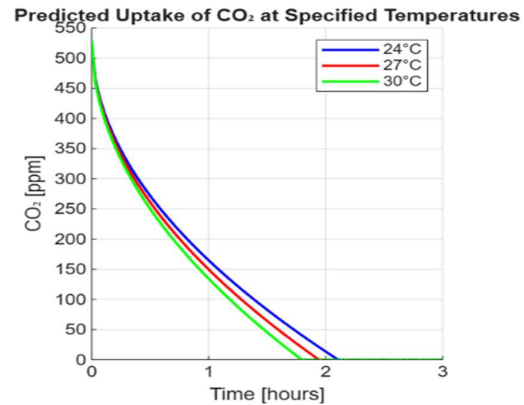
Despite these differences in early-age kinetics, the total CO₂ uptake remained within experimental uncertainty for all three temperatures. Thus, under the conditions tested (high relative humidity and short curing duration), temperature primarily affected the rate at which the steady-state value was approached, rather than the overall uptake over 24 hours.

**Figure 7.** Comparison of CO₂ concentration trends across curing temperatures.

4.4 Carbonation prediction model for varying temperatures

The equations derived in Section 2.1 were used to create a MATLAB model to predict the effect of

temperature on coke-modified concrete. Information such as cement capacity (amount of cement per m³ of concrete) was calculated based on our concrete compositions (seen in Table 1). The model assumes a perfectly sealed glove box. The model was used to estimate the diffusivity coefficient of CO₂ in coke-modified concrete. The effective diffusion coefficient was taken to be 1 x10⁻⁹ m²/s, slightly increased from literature values due to the increased porosity that the added coke provides.

**Figure 8.** Model prediction for CO₂ depletion in modified coke concrete at the three temperatures used in this project.

The activation energy was taken to be 20 kJ/mol, as taken from literature [24]. It is reasonably assumed that the activation energy will not vary a lot from normal concrete mix values. $F_c(W_c)$ was taken to be 1.2, an increase from the value of 1 used for normal concrete mixes. This slight increase accounted for increased porosity. Figure 8 displays the model's result, showing a rapid depletion of CO₂ after approximately 2 hours. The model deviated from the experimental results: it predicts faster carbonation at higher temperatures, whereas the measured initial rate decreased with temperature.

4.5 Uncertainty analysis

The duplicate trials demonstrated strong reproducibility, with only minor variation between experiments, as indicated by a standard deviation of less than 10 ppm. Small fluctuations were observed in all temperature conditions, particularly in the low-concentration region where sensor noise becomes more pronounced. Measurements conducted at higher temperatures showed increased noise levels. This is consistent with known NDIR sensor sensitivity to moisture and thermal drift [25]. For the low-concentration region, the noise standard deviation was approximately 10 ppm, with a marginal increase in the 30 °C experiment.

Table 5. Standard deviation for noise.

Temperature	24°C	27°C	30°C
σ_{final} (ppm)	5.656	6.903	14.940

To quantify background CO₂ leakage into the glove box, the concrete sample was removed while preventing fresh air from entering.

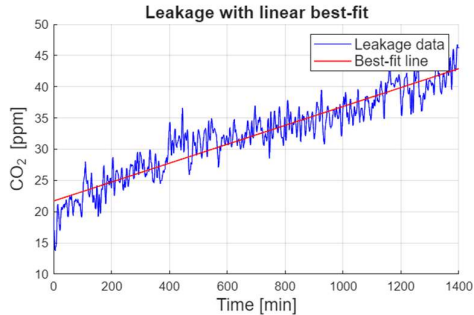


Figure 9. Line of best fit for leakage rate.

Under these conditions, the CO₂ concentration displayed a linear increase of 0.0151 ppm·min⁻¹. This leakage rate was used to correct all absorption data by subtracting it from the measured CO₂ decrease during the curing experiments. The equation for the leakage correction becomes:

$$C_{corrected}(t) = C_{measured}(t) - (0.0151 \times t) \quad (23)$$

Accounting for sensor accuracy and replicate variability, the total uncertainty in the calculated CO₂ concentration was estimated to be ± 18 ppm.

5. Discussion

This study examined the impact of temperature on the initial carbonation of coke-modified concrete during the first 24 hours of curing in a controlled environment. The addition of coke increased the initial CO₂ absorption rate compared to a control mix but did not significantly change the total uptake over 24 hours. Within the narrow temperature range studied (24–30 °C) and at high relative humidity, higher temperatures led to slower apparent early-age carbonation, in contrast to simple Arrhenius-type expectations.

5.1 Early-age carbonation behaviour

The strong front-loading of CO₂ uptake in the first 2–4 h is consistent with carbonation being initially controlled by rapid reaction with readily available pore sites and high gas–solid contact in partially saturated pores. This behaviour aligns with models that predict a moving carbonation front and decreasing effective diffusivity as calcium carbonate precipitates and fills pores [26].

The near-constant CO₂ concentration beyond t_{final} indicates that the system approaches a quasi-equilibrium within one day for the tested conditions. Longer-term measurements would be required to capture the slower carbonation of deeper regions.

5.2 Role of coke addition

The higher initial uptake rate for coke-modified concrete suggests that replacing a fraction of sand with coke enhances initial gas transport and reaction. The porous, hydrophilic coke phase likely increases the surface area and helps retain pore water, improving CO₂ dissolution and reaction with concrete. The similar total uptake after 24 hours for both mixes indicates that coke mainly accelerates the kinetics rather than significantly increasing short-term capacity under the conditions tested.

5.3 Temperature effects and mechanism

Contrary to an Arrhenius model expectation that higher temperatures should accelerate carbonation, the experiments showed slower initial uptake at 30 °C than at 24 °C. A likely explanation is the dependence of pore moisture and CO₂ solubility on temperature: at high relative humidity, increasing temperature can reduce CO₂ solubility in the pore solution, which may outweigh the acceleration of reaction kinetics.

Additionally, the short 24-hour observation window and thin sample geometry result in a measured rate that primarily reflects near-surface processes. Under these conditions, small changes in surface moisture level with temperature can dominate the response. A broader temperature and humidity range, tighter, and longer curing times would be required to separate diffusion-controlled from reaction-controlled regimes and observe long-term absorption behaviour.

5.4 Model interpretation

The experimentally observed decrease in initial carbonation rate with increasing temperature contrasts with the model prediction of a monotonic increase, and this discrepancy arises from simplifying assumptions in the current formulation. In the model, temperature only appears through an Arrhenius factor on the effective CO₂ diffusivity, while the surface CO₂ concentration, pore moisture state, and reaction kinetics are all treated as temperature-independent. Physically, however, raising the temperature at fixed relative humidity reduces pore water content and lowers CO₂ solubility in the pore solution, so the dissolved CO₂ available to react at the carbonation front can decrease even as molecular diffusivity increases. Over the narrow 24–30 °C range investigated, the experiments therefore likely reflect a regime where moisture and solubility effects dominate over the increase in diffusivity, making the negative temperature dependence more realistic than the current model response.

To better capture this behaviour, the model should be extended to include a temperature-dependent surface concentration via Henry's law, a moisture-dependent diffusivity linked to sorption isotherms (so that the diffusion coefficient varies with pore

hydration and with phase changes), and, if needed, separate activation energies for transport and reaction so that the effective rate is not constrained to increase with temperature.

5.5 Limitations and practical implications

The present work is limited to a narrow temperature window due to the limitations of the heater, high relative humidity, thin specimens, and a 24-hour curing period; therefore, the findings should not be directly extrapolated to long-term service conditions. Sensor accuracy, leakage correction and humidity fluctuations also introduced uncertainty that may disturb subtle temperature effects. Despite these limitations, the results indicate that coke addition can significantly accelerate early-age carbonation without reducing short-term uptake under the tested conditions, which could be advantageous for accelerated carbon-curing processes. Coke addition can be implemented when fast curing is required, without compromising the durability of the concrete [27] such as in highway, road and emergency infrastructure repairs.

To assess the practical viability of coke-modified concrete for large-scale CO₂ sequestration, future studies should explore a wider but realistic range of temperatures and humidities for curing conditions, using thicker concrete samples. Furthermore, the early-age kinetic measurements can be coupled with mechanical strength testing to estimate the performance of modified concrete. Finally, long-term carbonation depth measurements and life-cycle assessment of coke processing should be conducted to understand the environmental impacts of the process.

6. Conclusions

This study investigated the initial carbonation of control and coke-modified concrete at 24–30 °C using a sealed glove-box setup with continuous CO₂ monitoring. The addition of Coke increased the early CO₂ uptake rate relative to the control mix, while the total uptake over 24 hours remained similar. Within the tested temperature range, higher curing temperatures reduced the initial uptake rate but had little effect on the total 24-hour uptake.

These findings suggest that coke modification can accelerate early-age CO₂ sequestration in concrete without compromising short-term uptake capacity, which may be advantageous for processes that rely on rapid carbonation during curing. The weak temperature sensitivity of total 24 h uptake indicates that, under high humidity, modest variations in curing temperature may primarily affect the timescale rather than the extent of early carbonation.

The work is limited by the short curing period, thin specimens, restricted temperature range and high relative humidity, as well as uncertainties associated

with sensor accuracy and leakage correction. Future studies should extend the curing duration, explore wider temperature and humidity ranges, and couple experiments with more detailed transport–reaction modelling to evaluate the long-term performance and industrial feasibility of coke-modified concrete for CO₂ sequestration.

Acknowledgements

We would like to extend our thanks and gratitude to MSc student Hessa Alhashmi, as well as PhD students John Siwek, Aaron McQuade, Matthew Shields, and Dimitrios Nikolis, for their continuous support throughout this project.

References

- [1] ‘Causes of climate change - Climate Action - European Commission’. Accessed: Dec. 11, 2025. [Online]. Available: https://climate.ec.europa.eu/climate-change/causes-climate-change_en
- [2] ‘Visualized: Global Carbon Emissions by Sector’. Accessed: Dec. 11, 2025. [Online]. Available: <https://www.visualcapitalist.com/visualized-global-carbon-emissions-by-sector/>
- [3] ‘ICE response to the Spring Statement | Institution of Civil Engineers (ICE)’. Accessed: Dec. 11, 2025. [Online]. Available: <https://www.ice.org.uk/news-views-insights/media-centre/press-releases-list/ice-response-to-the-spring-statement>
- [4] ‘Cement Production by Country 2025’. Accessed: Dec. 11, 2025. [Online]. Available: <https://worldpopulationreview.com/country-rankings/cement-production-by-country>
- [5] ‘Construction building materials: commentary July 2025’, GOV.UK. Accessed: Dec. 11, 2025. [Online]. Available: <https://www.gov.uk/government/statistics/building-materials-and-components-statistics-july-2025/construction-building-materials-commentary-july-2025>
- [6] ‘UK concrete and cement industries slash emissions but risk of offshoring is growing, warns MPA’. Accessed: Dec. 11, 2025. [Online]. Available: <https://www.mineralproducts.org/News/2025/release29.aspx>
- [7] ‘Cement Manufacturing: Ways to Reduce CO₂ Emissions’. Accessed: Dec. 11, 2025. [Online]. Available: <https://www.ennomotive.com/cement-manufacturing-emissions>
- [8] ‘Global cement production has plateaued over the last decade - Our World in Data’. Accessed: Dec. 11, 2025. [Online]. Available: <https://ourworldindata.org/data->

- insights/global-cement-production-has-plateaued-over-the-last-decade
- [9] S. F. Arnold, Y. Shao, and G. Ramachandran, 'Evaluating well-mixed room and near-field-far-field model performance under highly controlled conditions', *J Occup Environ Hyg*, vol. 14, no. 6, pp. 427–437, June 2017, doi: 10.1080/15459624.2017.1285492.
 - [10] D. Li, L. Li, and X. Wang, 'Mathematical modelling of concrete carbonation with moving boundary', *International Communications in Heat and Mass Transfer*, vol. 117, p. 104809, Oct. 2020, doi: 10.1016/j.icheatmasstransfer.2020.104809.
 - [11] A. AL-Ameeri, M. I. Rafiq, O. Tsioulou, and O. Rybdylova, 'Impact of climate change on the carbonation in concrete due to Carbon Dioxide ingress: Experimental Investigation and Modelling', *Journal of Building Engineering*, vol. 44, p. 102594, May 2021, doi: 10.1016/j.jobe.2021.102594.
 - [12] H. A. Y. Alshaer *et al.*, 'Review on Carbonation Study of Reinforcement Concrete Incorporating with Bacteria as Self-Healing Approach', *Materials (Basel)*, vol. 15, no. 16, p. 5543, Aug. 2022, doi: 10.3390/ma15165543.
 - [13] C. Cheah and M. Ramli, 'Optimization of Mix Proportion of High Performance Mortar for Structural Applications', *American Journal of Engineering and Applied Sciences*, vol. 3, Apr. 2010, doi: 10.3844/ajeassp.2010.643.649.
 - [14] H. Mehdizadeh, X. Jia, K. H. Mo, and T.-C. Ling, 'Effect of water-to-cement ratio induced hydration on the accelerated carbonation of cement pastes', *Environmental Pollution*, vol. 280, p. 116914, July 2021, doi: 10.1016/j.envpol.2021.116914.
 - [15] J. M. Marangu, J. K. Thiong'o, and J. M. Wachira, 'Review of Carbonation Resistance in Hydrated Cement Based Materials', *Journal of Chemistry*, vol. 2019, no. 1, p. 8489671, 2019, doi: 10.1155/2019/8489671.
 - [16] B. Jaworska, D. Stańczak, R. Kobylka, T. Gołofit, D. Zhang, and J. Kuziak, 'The Influence of Fly Ash Carbonation on the Protective Properties of Concrete Cover Towards Reinforcement', *Materials*, vol. 18, no. 10, p. 2181, Jan. 2025, doi: 10.3390/ma18102181.
 - [17] L. Nayak, M. Rahaman, and R. Giri, 'Surface Modification/Functionalization of Carbon Materials by Different Techniques: An Overview', 2019, pp. 65–98. doi: 10.1007/978-981-13-2688-2_2.
 - [18] D. Fischbach, 'Kinetics of Graphitization—I. The High Temperature Structural Transformation in Pyrolytic Carbons'. Feb. 01, 1966.
 - [19] Y. Pu *et al.*, 'Accelerated carbonation technology for enhanced treatment of recycled concrete aggregates: A state-of-the-art review', *Construction and Building Materials*, vol. 282, p. 122671, May 2021, doi: 10.1016/j.conbuildmat.2021.122671.
 - [20] Z. Xu *et al.*, 'Effects of temperature, humidity and CO2 concentration on carbonation of cement-based materials: A review', *Construction and Building Materials*, vol. 346, p. 128399, Sept. 2022, doi: 10.1016/j.conbuildmat.2022.128399.
 - [21] I.-S. Yoon, O. Çopuroğlu, and K.-B. Park, 'Effect of global climatic change on carbonation progress of concrete', *Atmospheric Environment*, vol. 41, no. 34, pp. 7274–7285, Nov. 2007, doi: 10.1016/j.atmosenv.2007.05.028.
 - [22] N. Venkat Rao and T. Meena, 'A review on carbonation study in concrete', *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 263, no. 3, p. 032011, Nov. 2017, doi: 10.1088/1757-899X/263/3/032011.
 - [23] H. Cui, W. Tang, W. Liu, Z. Dong, and F. Xing, 'Experimental study on effects of CO2 concentrations on concrete carbonation and diffusion mechanisms', *Construction and Building Materials*, vol. 93, pp. 522–527, Sept. 2015, doi: 10.1016/j.conbuildmat.2015.06.007.
 - [24] G. Li, Y. Yuan, J. Du, and Y. Ji, 'Determination of the apparent activation energy of concrete carbonation', *Journal of Wuhan University of Technology Materials Science Edition*, vol. 28, no. 5, pp. 944–949, Oct. 2013, doi: 10.1007/s11595-013-0798-y.
 - [25] J. Kim and S. Yi, 'Effects of Temperature and Humidity on NDIR CO₂ Gas Sensor', *Journal of Sensor Science and Technology*, vol. 26, no. 3, pp. 179–185, 2017, doi: 10.5369/JSST.2017.26.3.179.
 - [26] A. Muntean, M. Böhm, and J. Kropp, 'Moving carbonation fronts in concrete: A moving-sharp-interface approach', *Chemical Engineering Science*, vol. 66, no. 3, pp. 538–547, Feb. 2011, doi: 10.1016/j.ces.2010.11.011.
 - [27] M. Rezaia, M. Panahandeh, N. Razavi, and F. Berto, 'Experimental study of the simultaneous effect of nano-silica and nano-carbon black on permeability and mechanical properties of the concrete', *Theoretical and Applied Fracture Mechanics*, vol. 104, p. 102391, Dec. 2019, doi: 10.1016/j.tafmec.2019.102391.

Superhydrophobic Modification of Different Cotton Types for Oil-Water Separation

Daniel Bull and Tony Na

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Many industries, such as mining, textiles, and metal processing, generate huge amounts of oily wastewater each year, posing serious economic and environmental threats. Since conventional oil-water separation techniques are often inefficient, expensive, and harmful to the environment, superhydrophobic cotton has emerged as a low-cost, sustainable option. This study compares the performance of four cotton types (cotton wool, synthetic, Egyptian, recycled) after superhydrophobic modification to identify the best candidate for industrial oil-water separation. Superhydrophobic cotton was fabricated via pre-treatment with NaOH, followed by surface modification with tetraethoxysilane (TEOS) and octadecylsilane (ODTMS). Notably, synthetic was the only cotton type that did not display superhydrophobic behaviour post-modification, as evidenced by its poor surface coverage in SEM images and high roll-off angle (24.6°). After a series of characterisation and performance tests on each modified cotton type, cotton wool was identified as the best performer. It exhibited the lowest roll-off angle (5.5°) and the lowest water sorption (0.46 g/g), indicating strong superhydrophobic behaviour. With the highest oil-water separation efficiency (98.6%), the cotton also demonstrated an exceptional ability to selectively absorb oil over water. Furthermore, it retained a low water sorption when exposed to harsh chemical conditions, showing its high durability. Modified Egyptian cotton also emerged as a promising candidate, displaying a low roll-off angle (7.2°), low water sorption (0.66 g/g), near-perfect oil-water separation efficiency (98.0%), and a high chemical durability across all media except detergent solution. In summary, superhydrophobic cotton wool and Egyptian cotton can provide a promising alternative to current oily wastewater treatment methods.

Keywords: oily wastewater, cotton, superhydrophobic, sorption tests, oil-water separation, durability

1. Introduction

In 2025, a massive fatberg was removed from a west London sewer since it was first discovered in 2017. The congealed lump of fat, grease, and non-biodegradable wastes threatened to block sewers and cause an overflow of wastewater with severe ecological and human health implications. While global oily wastewater discharge is rapidly increasing due to aging oil fields, industrial growth, and population increase [1][2], the fatberg incident urges the implementation of efficient and cost-effective oily wastewater treatment techniques. Conventionally, oily wastewater treatment relied on techniques such as gravitational, floatation, and evaporation oil-water separation methods. However, these methods either exhibit poor oil-water separation efficiency when oil is emulsified or require a high operating and capital expenditure [3]. Furthermore, life cycle analysis shows that traditional methods have directly impacted the environment through the extensive use of fossil fuels as an energy source, emitting greenhouse gases with high global warming potential [4]. Therefore, to meet the demanding effluent discharge requirements in a sustainable and cost-efficient way, superhydrophobic, oleophilic membrane processes have emerged as a promising candidate for oil-water separation.

Due to its simplicity, high separation efficiency, and low chemical demand, membrane separation has gained remarkable attention [5]. Superhydrophobic materials have become the research focus for oleophilic membrane separation due to their exceptionally high oil-water separation efficiency. If the membrane is coated with a superhydrophobic substance, the membrane can strongly repel water molecules while selectively absorbing oil from the mixture.

A superhydrophobic surface is defined as having a static water contact angle (WCA) of more than 150° and roll-off angles below 10° [6]. In nature, Lotus leaves exhibit an exceptionally high WCA, with water forming beads on the surface due to their double-scale roughness and epicuticular wax layer composed of long hydrocarbon chains. By exploiting these characteristics, materials such as stainless steel, sponge, plastic, and cotton have been modified to achieve superhydrophobicity suitable for oil-water separation by altering surface roughness and chemistry [7]. This research focuses on modifying cotton as a medium for oil-water separation for the following reasons. Cotton is a naturally abundant and biodegradable material, making it an environmentally sustainable resource. In addition, cotton fibre is primarily composed of a natural polymer, cellulose. The abundant hydroxyl functional groups in cellulose can facilitate the modification process, which is further discussed in this study. Finally, their porous structure can exhibit

a large oil sorption capacity per gram of cotton, which is advantageous for separation efficiency and cost-effectiveness for industrial oily wastewater treatment.

Conventional techniques for superhydrophobic surface modification rely on methods such as laser-by-laser assembly, laser process, sol-gel technique, and Hummers' method [8]. However, the listed methods typically require specialised equipment, complex reactions, and a high amount of energy, which limits their scalability for wider use. This research focuses on modifying cotton using a one-pot solution-immersion method, where the material is coated with a hydrophobic layer by chemical adhesion. The simplicity of the method allows it to be both cost-effective and scalable. A one-pot method involving the co-hydrolysis and co-condensation of tetraethyl orthosilicate (TEOS) and octadecyltrimethoxysilane (ODTMS) was selected for this study.

The main objective of this research was to fabricate a sustainable and high-performing superhydrophobic cotton as a solution for oily wastewater treatment. In addition, four different cotton types (cotton wool, Egyptian cotton, synthetic cotton, recycled cotton) were modified to compare and investigate the effect of distinct cotton properties on superhydrophobic performance and characterisation. Superhydrophobicity was characterised using SEM image analysis, roll-off angle measurements, and water sorption tests, while separation performance was evaluated through oil sorption, oil-water separation efficiency, and durability testing.

2. Materials and Methodology

Unless stated otherwise, all measurements were performed on four cotton types (cotton wool, synthetic, Egyptian, and recycled), prepared under identical conditions.

2.1 Materials

Cotton wool and TEOS ($\geq 99.0\%$) were purchased from Sigma Aldrich (Gillingham, UK). 100% PET polymer synthetic cotton was purchased from Homewit (Cardiff, UK), 100% wild Egyptian cotton was purchased from World of Wool (West Yorkshire, UK), and 100% recycled cotton was purchased from FibreLab (London, UK). Ethanol ($\geq 99.8\%$), sodium hydroxide (NaOH) pellets, acetone ($\geq 99.8\%$), hydrochloric acid (HCl) (37%), ammonia solution ($\geq 28\%$), and ODTMS (90%) were purchased from VWR (Lutterworth, UK). Deionised (DI) water was supplied by PURELAB Chorus 1 (ELGA LabWater) water purification system. Cooking oils (olive, vegetable, sesame, rapeseed, and sunflower oil) were sourced from a local supermarket.

All chemicals and reagents were used without further purification.

2.2. Pre-treatment of cotton

The cotton sample was rinsed with acetone three times, then gently squeezed to remove excess acetone. This was followed by drying in a vacuum oven at 60 °C and 100 mb for 1h. The dried cotton was subsequently immersed in a 1 M NaOH solution for 15 minutes. After etching, the samples were rinsed with DI water five times before drying in the vacuum oven at 60 °C and 200 mb until completely dry.

2.3. Superhydrophobic modification of cotton

To functionalise the cotton with a silica coating, a solution of ethanol (1.7 mol, 100 mL), DI water (0.6 mol, 10 mL), TEOS (0.03 mol, 6 mL), and ODTMS (6 mmol, 3 mL) was prepared and mixed for 10 minutes at 300 rpm using a magnetic stirrer. 1.5g of pre-treated cotton was then fully submerged in the solution and allowed to soak for 5 min.

2.4. Characterisation of cotton samples

After coating with gold (15 nm), the surface morphology of cotton samples was inspected using a scanning electron microscope (SEM) at 5 kV, specifically a Leo Gemini 1525 (Carl Zeiss AG, Germany).

In this study, roll-off angles were measured as a proxy for surface wettability. For each test, modified cotton samples were affixed to glass slides using double-sided tape. Gentle manual pressure was applied to obtain a uniformly flat surface. A 20 μ L droplet of deionised water was gently dispensed onto the cotton surface using a micropipette. The glass slide supporting the sample was then elevated by placing additional slides beneath one side, increasing the tilt angle until the droplet started to roll. The resulting tilt angle was quantified by importing the captured image into angle-measurement software (Angle Meter), which was used to measure the angle.

2.5. Water sorption

Water sorption was measured for both modified and unmodified cotton by submerging the samples in DI water for 5 min. A small plastic rod was used for modified samples to keep the cotton submerged and prevent it from floating. The soaked cotton was then removed, drained on a filter to remove excess water, and subsequently weighed. The water sorption capacity, S_W , was calculated using the equation below:

$$S_w = \frac{m_f - m_i}{m_i} = \frac{m_{\text{water absorbed}} [g]}{m_{\text{cotton}} [g]} \quad (1)$$

where m_i and m_f denote the initial and final mass of the cotton respectively.

2.6. Oil sorption

Maximum oil sorption for each oil type was measured by immersing the modified cotton samples in the respective oil for 60 min. The soaked cotton samples were then removed, put on a steel mesh to drain excess oil, and subsequently weighed. The oil sorption capacity, S_o , was calculated using the equation below:

$$S_o = \frac{m_f - m_i}{m_i} = \frac{m_{\text{oil absorbed}} [g]}{m_{\text{cotton}} [g]} \quad (2)$$

where m_i and m_f denote the initial and final mass of the cotton, respectively.

2.7. Oil-water separation

2.7.1. Batch experiments

For the batch experiments, 1 mL of olive oil was added to 9 mL of DI water in a measuring cylinder, followed by 100 mg of modified cotton. The system was left for 5 minutes to allow oil uptake. The soaked cotton was then removed and weighed. A challenge in accurately determining the mass of oil absorbed by the cotton was the presence of simultaneously absorbed water. To adjust this, the volume of residual water was measured directly from the cylindrical tube. The mass of water absorbed by the cotton was then subtracted to isolate the mass of oil uptake. The oil-water separation efficiency (η) was calculated using the equations below:

$$\eta (\%) = \frac{m_{o,abs}}{m_{o,initial}} \times 100 \quad (3)$$

where $m_{o,initial}$ is the initial mass of oil in the oil-water mixture, and $m_{o,abs}$ is the mass of oil absorbed by the cotton.

2.7.2. Flow system experiments

For the flow system experiments, a bench-scale packed column setup was used. A syringe was packed with 150 mg of modified Egyptian cotton and voids within the bed were minimised. An oil-water mixture was prepared by adding 8 mL of olive oil to 400 mL of DI water and mixing the phases

using a magnetic stirrer at 500 rpm. The packed syringe was connected to the beaker containing this mixture via a peristaltic pump, which was operated at 0.6 rpm (0.09 mL/s). Glass vials were placed beneath the setup to collect the effluent at 2-minute intervals for unsaturated cotton, 30-second intervals for partially saturated cotton, and 10-second intervals for fully saturated cotton. Each vial contained 54 mL of effluent. Following oil extraction with hexane, a UV-Vis spectrometer (nanodrop 2000c, Thermo-Fisher Scientific Inc., USA) was subsequently used to quantify the oil content in each vial.

2.8. Durability tests

Chemical durability was assessed by submerging ~25 mg of modified cotton in five test media for 24h: 0.5 wt% detergent solution, NaOH solution (pH 12), HCl solution (pH 1), acetone, and hot DI water at 65 °C. The samples were then removed, rinsed three times with ethanol, and dried in a vacuum oven at 60 °C and 100 mb for 1h. Water sorption tests were subsequently performed on the treated samples.

3. Results and Discussion

3.1. Superhydrophobic synthesis process

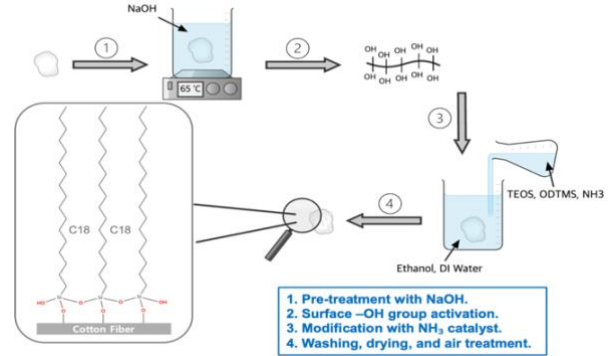


Figure 3.1.1 – A schematic representation of the superhydrophobic cotton synthesis process.

Figure 3.1.1 illustrates a schematic representation of the synthesis process. The process is composed of two steps: (1) the pre-treatment step and (2) the modification process. The pre-treatment step involves submerging the pristine cotton in 1M NaOH solution at 65 °C for 10 minutes. This chemical etching with NaOH increases the surface roughness of the cotton, as the alkaline solution removes materials of low molecular weight such as the wax layer and lignin on the surface [9]. The removal of surface impurities exposes hydroxyl ($-OH$) functional groups on the cellulose surface. However, previous research demonstrated that employing a highly concentrated alkaline solution caused a decrease in liquid sorption capacity in cotton due to clumping of cotton fibres [10].

Therefore, 1.0 M of NaOH solution was used in this study.

Through the co-hydrolysis and co-condensation reaction of TEOS and ODTMS, silica particles are formed on the surface of the cotton fibre. NH_3 is used as a catalyst, which lowers the activation barrier for the hydrolysis of silane to silanol [11]. The cotton fibre's exposed hydroxyl functional group facilitates the condensation reaction between silanol ($Si - OH$) from the silica particle and the cotton fibre, forming a strong covalent bond. Moreover, the remaining two silanol groups from the silica particles can form siloxane bonds ($Si - O - Si$) with the adjacent silica particle [12], forming a stable hydrophobic coating as represented in the final stage of the schematic diagram.

The C18 alkyl chain from ODTMS gives rise to cotton's superhydrophobic behaviour. This is due to the absence of permanent dipoles so only weak London dispersion forces prevail, effectively lowering the surface energy of the cotton, γ_s [13]. According to Young's equation, this results in a higher WCA, meaning the surface becomes less wettable with water (*Appendix Table A1*) [14].

Moreover, if a Cassie Baxter state is assumed, as shown in *Figure 3.1.3* and *Figure 3.1.4b*, the WCA also rises with the increase of surface gas fraction (*Appendix Table A1*) [15], which can be achieved by adding surface roughness. Hence, surface chemistry and surface roughness are the two factors that influence the wettability of a surface.



Figure 3.1.2 – Cotton wool repelling blue dyed water, forming a bead of droplet.



Figure 3.1.3 – A plastron layer formed around the cotton, indicating the Cassie-Baxter state [15].

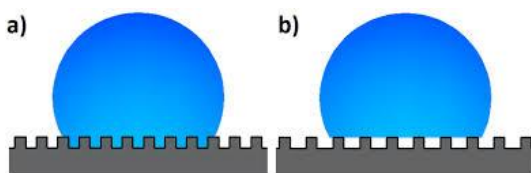


Figure 3.1.4 – A schematic representation of (a) Wenzel state and (b) Cassie-Baxter state. Wenzel state represents a weak hydrophobic state [16][17][18] where gas pockets are filled with liquid, whereas the Cassie-Baxter state represents a strong hydrophobic state due to a heterogeneous surface. Image courtesy of Darmanin, T., & Guittard, F (2015) [40].

3.2. Modified cotton morphology and wettability

Figure 3.2.1 and *Figure 3.2.2* represent SEM images of four different types of cottons before and

after modification. Prior to modification, natural cotton types (cotton wool, Egyptian, recycled) exhibited a subtly rough morphology as shown by the shallow grooves and ridges. This is because natural fibres are composed of bundled microfibrils [19]. In contrast, synthetic cotton exhibited a smooth surface morphology as PET fibre is a molecularly homogeneous substance.

Upon modification, cotton wool, Egyptian cotton, and recycled cotton exhibited a very rough coating of silica particles, with a high surface coverage of clusters of silica particles. Synthetic cotton, in contrast, showed scattered coverage with largely exposed, uncoated fibre areas. The difference may be due to different functional groups present in each of the polymers. Hydroxyl functional groups are abundant in a cellulose polymer, which induces a strong hydrogen bond between silica particle and cellulose. Polyethylene (PET) polymer, on the other hand, is composed of ester ($-COO$) groups, limiting the ability to form strong covalent bonds with the silica particles. Therefore, the silica layer adheres poorly to the fibre, resulting in a hydrophobic coating prone to removal.

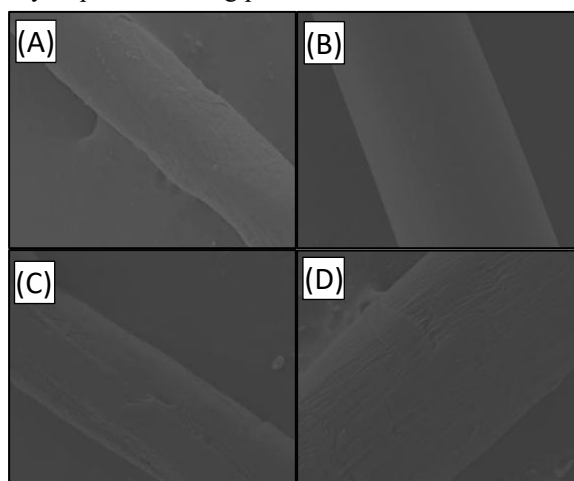


Figure 3.2.1 – SEM images of different types of unmodified cotton (A) Cotton Wool (B) Synthetic cotton (C) Egyptian cotton (D) Recycled Cotton.

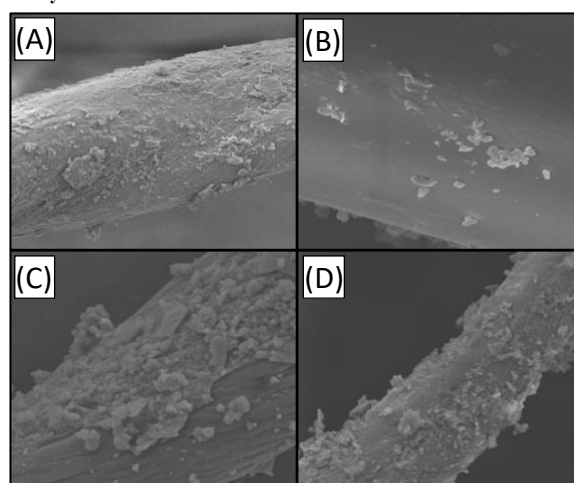


Figure 3.2.2 – SEM images of different types of modified cotton (A) Cotton Wool (B) Synthetic cotton (C) Egyptian cotton (D) Recycled Cotton.

The wettability of each modified cotton type was analysed using the roll-off angle and water sorption. Roll-off angle is typically used to characterise superhydrophobicity alongside WCA [20]. However, due to equipment failure, hydrophobicity was characterised by the roll-off angle and water sorption only.

Roll-off angle refers to the tilted angle of the surface when the stationary droplet starts to slide or roll due to gravitational force. Superhydrophobic surfaces are characterised by an extremely low roll-off angle, commonly $< 10^\circ$ [21].

As seen in Table 3.2.3, cotton wool, Egyptian cotton, and recycled cotton exhibited a very low roll-off angle below 10° , meaning these materials are likely to exhibit superhydrophobic behaviour. Synthetic cotton, on the other hand, required a higher roll-off angle, as the water droplet adhered to the cotton surface stronger than its peers. Previous research claims that materials with hydrophobic coating exhibited a roll-off angle of up to 90° [22], suggesting that synthetic cotton is more consistent with a hydrophobic regime rather than a superhydrophobic one.

	Cotton wool	Synthetic cotton	Egyptian cotton	Recycled cotton
Roll-off angle	5.5°	24.6°	7.2°	9.3°

Table 3.2.3 – Roll-off angles of various modified cotton types

3.3. Water sorption analysis

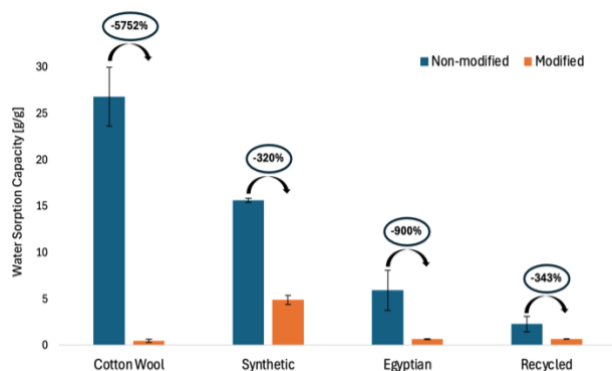


Figure 3.3.1 – Water sorption capacities of various modified and unmodified cotton types

For oily wastewater treatment, selectively absorbing oil and minimising water uptake is a highly desirable trait, motivating the analysis of water sorption capacities for each cotton type. Water sorption experiments were performed on both unmodified and modified samples to quantify the change in hydrophobicity for each cotton type.

As shown in Figure 3.3.1, unmodified cotton wool demonstrated the highest water sorption capacity of 26.7 g/g. Sample calculations can be found in Appendix D. Since cotton wool is often used for surgical and medical purposes, it is

extensively processed and purified to remove natural impurities from upland cotton. This includes scouring (alkaline boiling with NaOH) and bleaching with oxidising agents at high temperature and pressure [23][24]. These processes effectively remove surface impurities, leaving 99% cellulose chemical composition [25]. This industrial purifying process thus facilitates the silica particle coating on cotton wool.

In contrast, unmodified Egyptian and recycled cotton exhibited a low water sorption capacity of 5.9 and 2.3 g/g, respectively. This can be attributed to Egyptian cotton's low inter-fibre porosity and recycled cotton's low intra-fibre porosity (see Section 3.4).

Post-modification, synthetic cotton showed the highest water sorption capacity. This can be explained by synthetic cotton's surface chemistry, as discussed in Section 3.2, and bulk density. Figure 3.2.2 shows that the synthetic cotton surface remained relatively smooth post-modification. This is because the ester functional group does not form a strong covalent bond with the silanol group from the silica particles. Moreover, synthetic cotton exhibits low bulk density due to PET fibre's high crimp number. In the textile industry, crimp number refers to the frequency of waves in a staple [26]. PET fibres are artificially engineered to have 8–10 crimps per 25mm of fibre length [27] in a three-dimensional spiral form. This makes the fibres spring apart, resulting in a low bulk density, and high porosity.

3.4. Oil sorption capacity

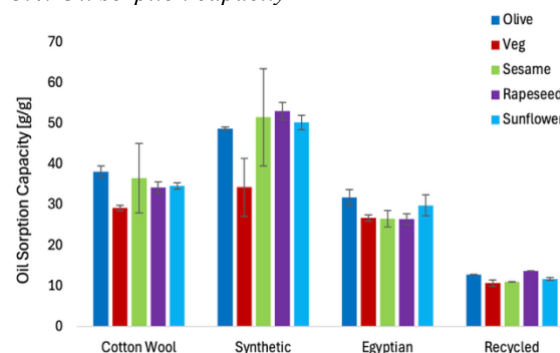


Figure 3.4.1 – Oil sorption capacity of different types of modified cotton after being fully submerged with 5 different types of oils (olive, vegetable, sesame, rapeseed, sunflower) for 60 minutes

For practical applications, a high oil sorption capacity is desirable for oily wastewater treatment. This is because it reduces the frequency of sorbent replacement, thereby lowering operating expenses and improving separation efficiency.

Across five oil types, synthetic cotton exhibited the highest oil sorption capacity (34.3–51.5 g/g), whereas recycled cotton exhibited the lowest (10.7–13.7 g/g).

The lower oil sorption capacity observed in recycled cotton can be attributed to its damaged fibre structure. Since oil absorption is also attributed to

the capillary force through the lumen (intra-fibre porosity), fibres with hollow, intact lumen typically exhibit a higher sorption capacity [28]. Recycled cotton is post-industrial waste obtained after mechanical shredding and milling. This can cause permanent damage to the structure of the cotton fibres, as shown by the disjointed fibres in *Figure 3.4.2*. Therefore, mechanical processing can collapse or damage the lumen structure, resulting in lower available pore volume for oil sorption (*Figure 3.4.3*) [29][30][39].

This result is in line with the previous research by Beom-gu Lee et al [31], which states that oil sorption is affected by grinding. Once the cotton fibres were ground to less than their original length, oil sorption capacity decreased.

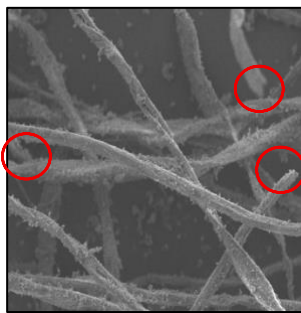


Figure 3.4.2 – SEM image of recycled cotton representing the disjointed and damaged cotton fibres/

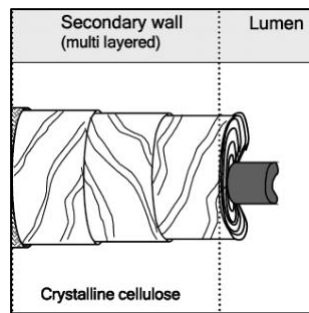


Figure 3.4.3 – Diagram representing the inner structure of the fibre. Image courtesy of Taylor and Francis [41].

All other cotton types were fully saturated after submerging in oil for 5 minutes, but Egyptian cotton exhibited a clear saturation gradient across the sample for each oil tested (*Figure 3.4.4*). The unsaturated fibre core shows that Egyptian cotton's sorption rate is significantly lower than its peers.

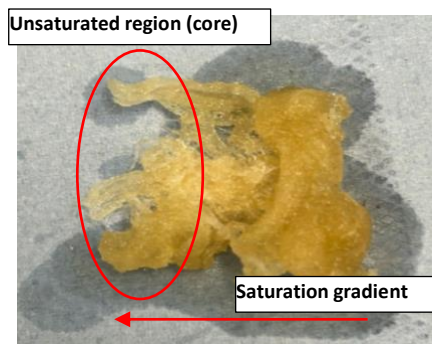


Figure 3.4.4 – Cross-section of Egyptian cotton showing an unsaturated cotton core and a saturation gradient

The low oil sorption rate observed in Egyptian cotton can be attributed to its cotton morphology. *Appendix Table A1* shows the cotton fibre length and diameter, confirmed by SEM analysis. Egyptian cotton exhibits a long fibre length and the smallest fibre diameter. A previous study conducted by Raj & Sreenivasan confirmed that cottons with smaller fibre diameter (low micronaire value) exhibited low

air permeability, which is attributable to its high packing density [32]. Therefore, the finer the cotton fibre, the smaller the inter-capillary pore diameter. According to Washburn's equation in its differentiated form, the rate of wicking $\frac{dh}{dt}$ is dependent on the inter-capillary pore radius r , surface tension of the liquid γ , liquid contact angle θ , liquid viscosity η , and the pre-wicked length l [33]:

$$\frac{dh}{dt} = \frac{r \cdot \gamma \cdot \cos \theta}{4 \cdot \eta \cdot l} \quad (4)$$

Therefore, Egyptian cotton's small fibre diameter decreases the overall liquid penetration rate within the cotton. This explains the observed saturation gradient inside Egyptian cotton when cut in half (*Figure 3.4.4*). 5 minutes was not sufficient for oil to infiltrate the small inter-capillary pores.

To account for this, Egyptian cotton was submerged in oil for 60 minutes to identify its true oil capacity. The sorbent's maximum oil sorption capacity (*Figure 3.4.5*) was substantially higher than the uptake measured after only 5 minutes of soaking.

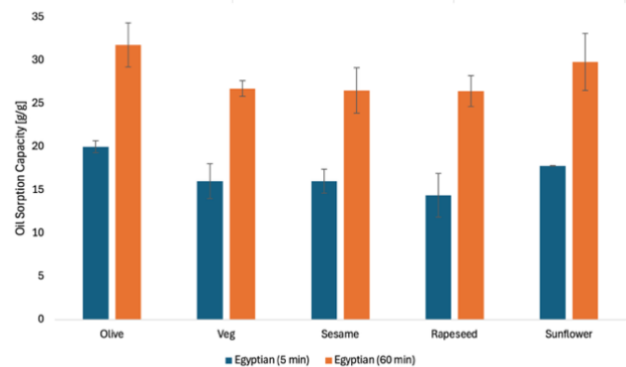


Figure 3.4.5 – Oil sorption capacity of Egyptian cotton submerged in 5 types of oils for 5 minutes (blue) and 1 hour (orange).

3.5. Oil-water separation

Research into superhydrophobic cotton as a candidate for oily wastewater treatment has caught considerable attention in modern times [34]. This section explored the oil-water separation performance of each modified cotton type in a batch system and then analysed the top-performing cotton in a continuous flow system.

3.5.1 Batch experiments

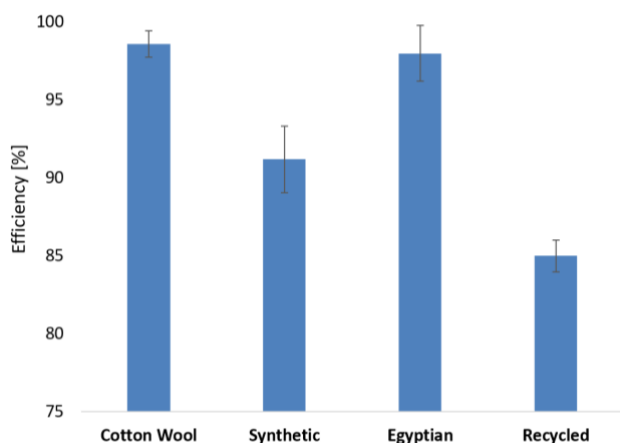


Figure 3.5.1 – Olive oil-water separation efficiency of different modified cottons in a batch system.

The experiment was straightforward to perform because olive oil and DI water formed clearly separated phases, with the oil floating due to its lower density. A low volume of oil (8 mL) was selected to prevent the cotton from becoming saturated.

As shown in Figure 3.5.1, there are clear differences in oil-water separation efficiencies between the four modified cotton types. Cotton wool and Egyptian cotton exhibited near-perfect separation efficiencies at 98.6% and 98.0% respectively, suggesting both oleophilic and superhydrophobic behaviour. Synthetic and recycled cotton followed with lower efficiencies at 91.2% and 85.0% respectively. Sample calculations can be found in Appendix D.

The higher efficiency of cotton wool and Egyptian cotton is consistent with their low water sorptions (Figure 3.3.1) and high maximum oil sorption capacities (Figure 3.4.1).

The lower efficiency of synthetic cotton can be attributed to its high water sorption, despite having the highest oil sorption capacity. As discussed in Section 3.2, this is due to scattered silica coating giving rise to exposed wettable regions, allowing greater water penetration and uptake.

As discussed in Section 3.4, recycled cotton having the lowest efficiency can be explained by its damaged fibre structure with collapsed lumens.

3.5.2 Flow system experiments

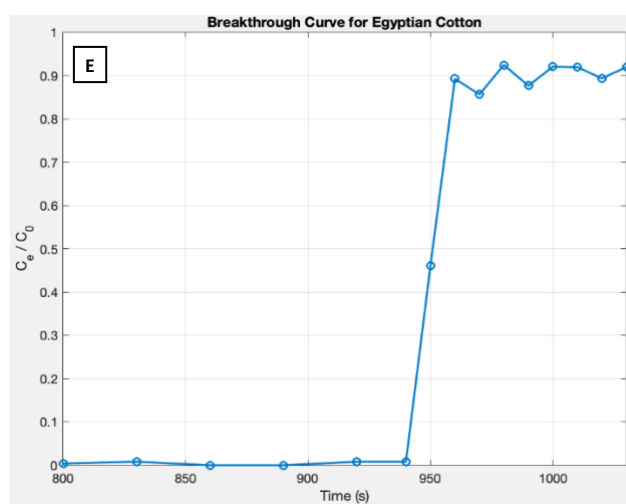
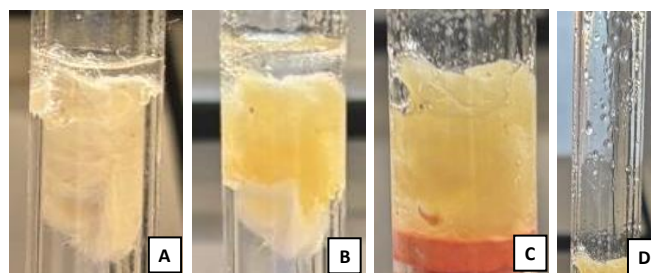


Figure 3.5.2 – (A) unsaturated Egyptian cotton packed in the syringe, (B) partially saturated Egyptian cotton packed in the syringe, (C) fully saturated Egyptian cotton packed in the syringe, (D) oil residue on syringe wall, (E) oil breakthrough curve for modified Egyptian cotton.

The aim of this experiment was to evaluate the oil-water separation performance of modified cotton in an industrial environment. To avoid repeating established flow-system findings on cotton wool, the next-best performer, Egyptian cotton, was selected for the analysis.

Figures 3.5.2A-C show the state of the cotton bed throughout the experiment (unsaturated, partially saturated, fully saturated). To avoid unnecessary use of consumables, the effluent was initially collected at 2-minute intervals for unsaturated cotton (Figure 3.5.2A) while the effluent was pure water. Samples were then collected at 30-second intervals for partially saturated cotton (Figure 3.5.2B) and 10-second intervals for fully saturated cotton (Figure 3.5.2C) to accurately measure the change in effluent concentration near breakthrough. The breakthrough curve was generated by plotting C_e/C_0 against time, where C_e denotes the effluent oil concentration and C_0 denotes the initial oil concentration of the mixture.

Figure 3.5.2E shows that until cotton saturation, very little oil was observed in the effluent. This suggests that the cotton bed was acting as an effective filter, preferentially absorbing oil at near-100% selectivity and allowing water to pass through. This is consistent with previous literature, where it was found that hydrophobic cotton could

selectively absorb oil from oily water with a nearly pure effluent of water [35]. The trace oil concentrations (e.g. at 830 seconds) can be put down to tiny amounts of oil in oil-water emulsions bypassing the cotton through inter-fibre voids.

After cotton saturation at the breakthrough point (~945 seconds), there was a significant jump in effluent oil concentration, with C_e/C_0 increasing to 0.89. Contrary to the typical S-shaped breakthrough curve, the increase observed was notably sharp, which represents the cotton's high adsorption kinetics [36].

Post-breakthrough, the ratio stayed relatively constant. The small variations in C_e/C_0 likely reflect imperfect mixing, leading to different local oil concentrations throughout the system. After the cotton becomes saturated, C_e/C_0 should theoretically equal unity to show that input and output oil concentrations are the same. The slightly lower value in practice can be attributed to the oil residue on the syringe walls, as observed in *Figure 3.5.2D*, as well as minor losses within the pump.

3.6. Durability tests for each modified cotton type

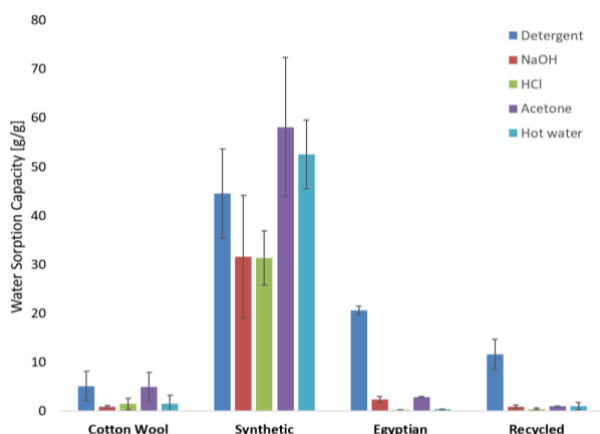


Figure 3.6.1 – Water sorption of modified cotton types after 24h immersion in durability media.

Understanding whether the cotton maintains its superhydrophobicity under harsh chemical conditions is crucial in judging how viable it is for wastewater treatment. Therefore, a durability analysis for each cotton type is of upmost importance.

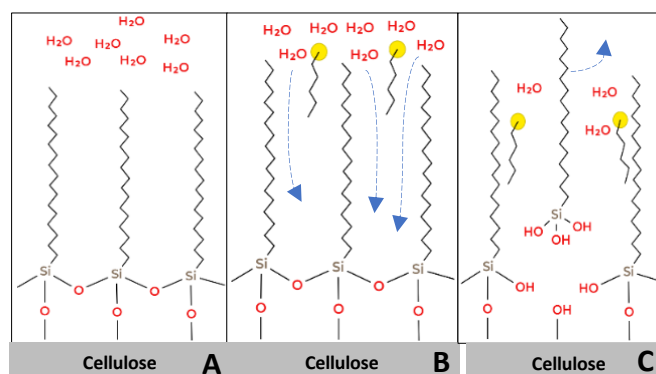
Chemical durability was evaluated by submerging the modified cotton in 0.5 wt% detergent solution, hot DI water at 65 °C, NaOH solution (pH 12), HCl solution (pH 1), and acetone for 24h. The water sorptions for each cotton type in every test media were then measured.

The water sorption data in *Figure 3.6.1* shows clear differences in chemical durability between the four modified cotton types. Synthetic cotton had by far the largest water sorption increase in every media, indicating the greatest loss in hydrophobicity. Cotton wool, Egyptian cotton, and

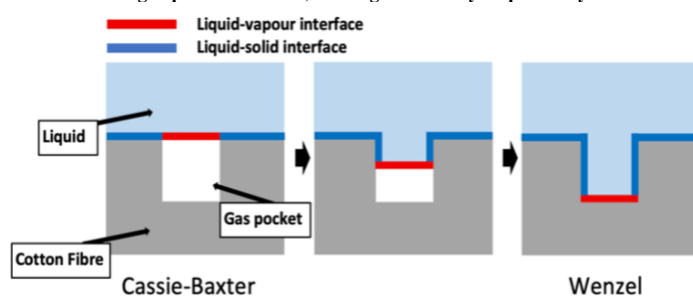
recycled cotton generally retained low water sorptions, with detergent being the exception (discussed below), suggesting higher durability overall. These results can be explained by coating chemistry, where the formation of strong Si–O–cellulose bonds in the hydrophobic layer (*Schematic 2 Step A*) depends on the availability of surface –OH groups [37]. As discussed in *Section 3.2*, synthetic cotton's hydrophobic coating is weakly bound and more easily disrupted by chemical exposure.

Detergent solution caused the largest increase in water sorption, with an increase of 39.7 g/g for synthetic cotton, 19.9 g/g for Egyptian cotton, 10.9 g/g for recycled cotton, and a small increase of 4.6 g/g for cotton wool. This suggests that modified cottons, except for cotton wool, were not durable under the detergent media.

Previous research indicates that theoretically, the Wenzel state is thermodynamically favourable to the Cassie-Baxter (CB) state at all surfactant concentrations. When a surfactant is introduced into a solvent, amphiphilic surfactants adsorb readily at the superhydrophobic surface and destroy the CB state by reducing surface tension, as illustrated in *Scheme 2*. As a result, gas pockets are filled with liquid [38]. This promotes water penetration, which then causes the partial removal of the coated layer as illustrated in *Scheme 3*.



Schematic 2 – Schematic illustration of how detergent surfactants cause the removal of superhydrophobic coating on modified cotton fibres. (A) Hydrophobic ODTMS layer with long C18 chains repels water. (B) Detergent surfactants (yellow circle = hydrophilic head, alkyl chain = hydrophobic tail) adsorb at the interface surface and increase wettability by filling in the air pockets (*Schematic 3*). This promotes water access to the interface, which can accelerate the penetration of water molecules inside the hydrophobic layer to attack the Si–O bonds. (C) Partial removal of coating exposes –OH sites, causing a loss of hydrophobicity.



Schematic 3 – Schematic illustration of detergent solution destroying the Cassie-Baxter state, resulting in the Wenzel state [40]

In summary, modified cellulose-based cottons (cotton wool, Egyptian cotton, recycled cotton) exhibited excellent durability, keeping their superhydrophobicity across most media. Detergent caused the largest increase in water sorption for these cottons, with Egyptian and recycled cotton showing a partial loss in hydrophobicity, and cotton wool retaining superhydrophobicity. Synthetic cotton performed the worst with a huge increase in water sorption in every medium, indicating a significant loss of superhydrophobicity.

4. Conclusion and Outlook

In conclusion, pre-treatment with NaOH followed by modification with TEOS and ODTMS synthesised cottons of varying superhydrophobicity. Notably, synthetic cotton was the only cotton type not to display superhydrophobic behaviour after modification, as shown by the scattered silica coating in SEM images and high roll-off angle (24.6°). Out of the four modified cotton types assessed, cotton wool was identified as the highest performer. It combined the lowest roll-off angle (5.5°) and water sorption (0.46 g/g) with high oil sorption capacity (29.2–38.1 g/g), the highest oil-water separation efficiency (98.6%), and the lowest water sorption after exposure to various durability media. Recycled cotton showed promising superhydrophobicity (roll-off angle 9.3°, water sorption 0.66 g/g) but was limited by its lower oil sorption capacity (10.7–13.7 g/g) and oil-water efficiency (85.0%). This was explained by its damaged fibre structure induced by recycling, which collapsed the lumens and reduced the available pore volume for oil sorption. Egyptian cotton also proved to be a promising candidate (roll-off angle 7.2°, water sorption 0.66 g/g, oil sorption 14.4–20.0 g/g, separation efficiency 98.0%), retaining low water sorption post-durability except in detergent. The oil-water separation performance of modified Egyptian cotton was therefore explored in a flow system. The breakthrough curve showed effectively zero effluent oil concentration until breakthrough, indicating near-100% selectivity, followed by a steep rise after cotton saturation. This highlights the importance of regeneration and sufficient packing in continuous operations to stop oil from escaping post-breakthrough. These collective findings show that superhydrophobic cotton wool and Egyptian cotton have high potential for use in oily wastewater treatment.

For future research, water contact angle (WCA) measurements could be conducted to confirm the cotton's hydrophobicity with greater confidence. X-ray photoelectron spectroscopy (XPS) can further aid surface characterisation by quantifying silicon content, offering an indirect measure of silica coverage. Moreover, a higher number of repetitions of experiments can be conducted to ensure the

reliability of the study. For instance, larger sample sizes in sorption tests could reduce standard deviation and enhance statistical confidence. Additionally, because emulsion formation represents the major challenge in industrial oil-water separation, evaluating the effect of different stirring speeds (RPMs) on separation performance would clarify whether the material can effectively treat emulsified mixtures across a range of shear conditions.

Overall, based on the preferred cotton characteristics as a medium for oil-water separation identified in this study, practical candidates for large-scale oil-water separation technologies can be further explored. For instance, potential candidates may include *G. arboreum* and *G. herbaceum*. These natural cottons inherently have a high cellulose composition, meaning they can be readily modified. Their short-stapled and coarse fibres may be advantageous for sorption rates. Finally, they exhibit low bulk density, which is attributable to their fibre stiffness, advantageous for oil sorption capacity.

Acknowledgements

The authors would like to thank Dr. Tsaone Gosiamemang, Dr. Isha Bade, and Prof. Jerry Heng for their invaluable guidance throughout the research.

Dilatational Surface Viscous Effects on Bubble Bursting at Low Bond Number

Hamza Al-Aedroos and Shreyas Sivakumar

Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom

Abstract

Three-dimensional numerical investigation of bubble bursting at a liquid interface saturated with an insoluble surfactant is conducted in this work. The interfacial dilatational Ohnesorge number is introduced (Oh_d), enabling the exclusive variation of the dilatational surface viscosity (μ_d) whilst keeping all other rheological and hydrodynamic parameters constant. Direct Numerical Simulations (DNS) are performed, solving the incompressible Navier-Stokes equation coupled with surface viscous stress using an advanced in-house multiphase framework, namely code BLUE [1]. Four cases are simulated, covering four orders of magnitude in dilatational surface viscosity to explore its effect on bubble bursting behaviour. Our results reveal that when $Oh_d < Oh$ (Oh refers to bulk Ohnesorge number), interfacial dynamics remain similar to those in the clean interface. However, pronounced deviation arises when Oh_d exceeds Oh , underlining the existence of distinct bulk viscous-dominated and surface dilatation-dominated regimes. Significant differences between the two regimes include (a) considerably hindered cavity collapse, (b) thicker, taller jet profiles and (c) larger ejected droplets.

Keywords: Aerosols, Contaminants, Surface rheology, Bubble bursting

1. Introduction

Each passing day, the ocean surface grows increasingly contaminated as anthropogenic waste accumulates at the air-sea interface. The accumulation of waste modifies the interfacial properties and alters the fluid mechanics of processes occurring there. Among the most consequential of these processes is bubble bursting [2] [3], the primary mechanism by which material is transferred into the atmosphere from the ocean's surface as aerosols. Aerosols act as dynamic energy moderators in the atmosphere, facilitating heat transfer from the ocean surface to the air, thereby contributing to global warming. Sea spray constitutes the largest mass flux of aerosols into the atmosphere, with an important impact on atmospheric radiative transfer [4]. This has attracted increasing attention regarding the underlying mechanism of bubble rupture and droplet formation. To understand how bursting bubbles generate aerosols, and to thereby understand

how droplet formation could be mitigated, it is necessary to examine the mechanisms and physical properties that govern their dynamics. As discussed thoroughly by Boulton-Stone et al [5] in their study on gas bubble bursting at a free surface, the rupture of a bubble during bursting is capillary-driven. The subsequent cavity collapse, excites capillary waves that are focused at the axis of symmetry to produce a Worthington jet [6] [7] [8]. Following this, spherical droplets of liquids break off from the jet due to Rayleigh-Plateau instability, the surface tension mechanism by which liquids minimise their surface area for a given volume [9].

Having established the fundamental capillary and inertial mechanisms that drive bubble bursting, the discussion now turns to how surface rheology alters the fluid mechanics of the system. At a clean surface, the dynamics of bubble bursting is determined by two dimensionless numbers: the Ohnesorge number $Oh = \mu / \sqrt{\rho \sigma R_0}$ which compares bulk viscous forces

to inertio-capillary forces, and the Bond number $Bo = \rho g R_0^2 / \sigma$ which compares gravity to capillarity. However, the assumption of a clean surface simply does not hold true in several applications due to the presence of surface-active contaminants (surfactants), such as biological agents, proteins, lipids and polymers. These surfactants adsorb at fluid-fluid interfaces and modify interfacial properties, giving rise to surface rheological effects.

The introduction of a surfactant alters the surface tension, generating surface elastic stresses tangential to the interface. On the other hand, the presence of these agents also gives rise to in-plane friction, resulting in surface rheological effects. Surface elasticity can be defined as the reversible change to surface tension during interfacial area changes. Surfactant concentrations along the interface generate surface tension gradients, which drive Marangoni stresses, the mechanism that drives fluid flow from areas of high to low surface tension. Surface viscosity is the resistance to flow in the plane of the interface, and can be broken down into shear and dilatational surface viscosities. Shear is the resistance to sideways deformation along the interface and dilatation is the resistance to interfacial area change during compression and expansion.

With regards to elasticity, the interfacial elasto-capillary number E_c is used to compare the effects of interfacial elasticity and capillarity. An increased value of E_c has been shown to effectively weaken droplet ejection by retarding capillary wave formation [10] [11]. Furthermore, the effects of Marangoni stresses on the dynamics of bursting bubbles are well documented by Constante-Amores and colleagues in their 2021 study [8]. They notably remarked that the presence of these stresses retard capillary wave propagation and subsequent jet evolution, thus limiting the number of ejected droplets. This is further substantiated by the investigations of Pico et al. [7]

Although surface elastic effects have been extensively investigated, research on surface viscosity remains scarce. From the limited work conducted, the literature clearly establishes that an increase in surface dilatational viscosity slows and dampens the formation of a liquid jet [12] [13]. In addition, Boulton-Stone, in his 1995 study on the effect of surfactant on bursting gas bubbles, highlights that

surface shear viscosity also has an effect, granted it is smaller in magnitude [14]. Therefore, it is clear that surface viscosity significantly alters interfacial flow patterns and behaviours. However, preceding studies implement two-dimensional models, which fail to capture the three-dimensional complexities of surface viscous effects. This, coupled with the fact that to the best of the authors' knowledge, no study has investigated bubble bursting mechanics with variations exclusively in dilatational surface viscosity, forms the motivation behind this study.

In this paper, we conduct four simulations, each modelling differing magnitudes of surface viscosity. A clean case with no surface viscous effects is used as a reference. We observe minor deviations as bulk viscous effects initially dominate. The transition into the surface viscous-dominated regime exhibits large deviations from the clean case: we observe a delayed profile that produces a thicker, taller jet, subsequently ejecting a droplet with a larger radius. In Section 2, we outline the governing equations and the DNS set up. In Section 3, we assess the impact of dilatational viscosity on the dynamics of bubble bursting. The paper then aims to explain the trends observed in Section 4, concluding with the potential implementations of droplet inhibition through insoluble surfactants in Section 5.

2. Problem Formulation

Table 1: Simulation Parameters.

Parameter	Symbol	Unit
Bubble Radius	R_0	m
Gravitational Constant	g	m/s^2
Liquid Density	ρ_l	kg/m^3
Liquid Viscosity	μ_l	Pa s
Surface Dilatational Viscosity	μ_d	kg/s
Surface Shear Viscosity	μ_s	kg/s
Gas Density	ρ_g	kg/m^3
Gas Viscosity	μ_g	Pa s
Surface Tension	σ	N/m

In this study, we conduct incompressible, three-dimensional direct numerical simulations for two-phase flows in a Cartesian coordinate system ($\mathbf{x} = (x, y, z)$) to investigate the role of surface dilatational viscosity on bubble rupture and jet

formation.

The initial domain setup for a clean interface is shown in Figure 1 and Table 1 contains the physical input parameters used in this study.

2.1. Governing Equations

The governing continuity and momentum (Navier-Stokes) equations can be written as:

$$\nabla \cdot \mathbf{u} = 0 \quad (1)$$

$$\rho \frac{D\mathbf{u}}{Dt} = -\nabla p + \nabla \cdot \mu (\nabla \mathbf{u} + \nabla \mathbf{u}^T) + \rho \mathbf{g} + \mathbf{f}(\mathbf{x}_f) \quad (2)$$

where, $D(\cdot)/Dt$ is the material derivative, $\mathbf{u}$ is the fluid velocity vector such that $\mathbf{u} = (u, v, w)$, p is the total pressure of the fluid and $\mathbf{g}$ is the gravitational acceleration vector such that $\mathbf{g} = (0, 0, -g)$. $\mathbf{f}(\mathbf{x}_f)$ represents the interfacial forces at the interface, $\mathbf{x} = \mathbf{x}_f$ and is defined as:

$$\mathbf{f}(\mathbf{x}_f) = \int_{A_s} \nabla_s \cdot \mathbf{T} \delta(\mathbf{x} - \mathbf{x}_s) dA_s \quad (3)$$

where

$$\mathbf{T} = \left[(\sigma + (\mu_d - \mu_s)(\nabla_s \cdot \mathbf{u})\mathbf{I}_s) + 2\mu_s \mathbf{D}_s \right] \quad (4)$$

Here $\mathbf{T}$ is the Boussinesq-Scriven stress tensor, $\mathbf{I}_s$ is the surface identity tensor $\mathbf{I}_s = \mathbf{I} - \mathbf{nn}$, where $\mathbf{n}$ is the unit normal vector to the interface and A_s is the interfacial surface area. The surface material properties are the surface tension (σ), dilatational surface viscosity (μ_d), and shear surface viscosity (μ_s). ρ and μ are the single-fluid density and viscosity respectively, defined by the Heaviside function, $\mathcal{H}(\mathbf{x}, t)$, which is 1 for the liquid phase and 0 for the gas phase. $\delta(\mathbf{x} - \mathbf{x}_f)$ is the Kronecker-delta function, which is non-zero if and only if, $\mathbf{x} = \mathbf{x}_f$. $\nabla_s(\nabla_s \cdot)$ is the surface gradient (divergence) operator and $\mathbf{D}_s$ is the surface deformation rate tensor, $\mathbf{D}_s = \frac{1}{2} [\nabla_s \mathbf{u}_s \cdot \mathbf{I}_s + \mathbf{I}_s \cdot (\nabla_s \mathbf{u}_s)^T]$.

2.2. Non-Dimensional Numbers

The numerical simulations are performed by solving the two-phase Navier-Stokes equations in a three-dimensional Cartesian domain as shown in Figure 1. A hybrid front-tracking-level-set method

known as the level contour reconstruction method (LCRM) [15] is implemented to formulate the equations presented in this paper. This allows accurate interfacial curvature (front-tracking) and topological changes during rupture to be accurately depicted. Physical characteristics are rendered dimensionless for the purpose of generalisation and scalability. The choice of non-dimensionalisation, as seen in Equations (5) and (6) is consistent with the preceding literature on bursting bubbles, such as the investigation conducted by Pico et al. [7]. Parameters imposed with a tilde are dimensionless and Table 2 outlines these quantities.

$$\tilde{\mathbf{x}} = \frac{\mathbf{x}}{R_0}, \quad \tilde{t} = \frac{t}{t_c}, \quad \tilde{\mathbf{u}} = \frac{\mathbf{u}}{u_c} \quad (5)$$

$$\tilde{p} = \frac{p}{\rho u_c^2}, \quad \tilde{\mu} = \frac{\mu}{\mu_l}, \quad \tilde{\rho} = \frac{\rho}{\rho_l}, \quad (6)$$

Table 2: Nondimensional Parameters.

Symbol	Meaning
$\tilde{\mathbf{x}}$	Spatial Vector
$\tilde{t}$	Time
$\tilde{\mathbf{u}}$	Velocity Vector
$\tilde{p}$	Pressure
$\tilde{\mu}$	Viscosity
$\tilde{\rho}$	Density
$\tilde{\sigma}$	Surface Tension

In Equation 5, t_c is the capillary timescale [s] where $t_c = R_0/u_c = \sqrt{\rho_l R_0^3/\sigma}$ and u_c is the velocity scale [m/s] where $u_c = \sqrt{\sigma/\rho_l R_0}$.

2.3. Non-Dimensionalised Equations

Both the liquid and gas phases are assumed to be Newtonian, hence they are immiscible and the flow is incompressible. The governing equations are made dimensionless using the scalings in Section 2.2 as follows:

$$\tilde{\nabla} \cdot \tilde{\mathbf{u}} = 0 \quad (7)$$

$$\tilde{\rho} \frac{D\tilde{\mathbf{u}}}{D\tilde{t}} = -\nabla \tilde{p} - Bo \hat{\mathbf{i}}_z + Oh \tilde{\nabla} \cdot [\tilde{\mu} (\tilde{\nabla} \tilde{\mathbf{u}} + (\tilde{\nabla} \tilde{\mathbf{u}})^T)] + \tilde{\mathbf{f}}(\tilde{\mathbf{x}}_f) \quad (8)$$

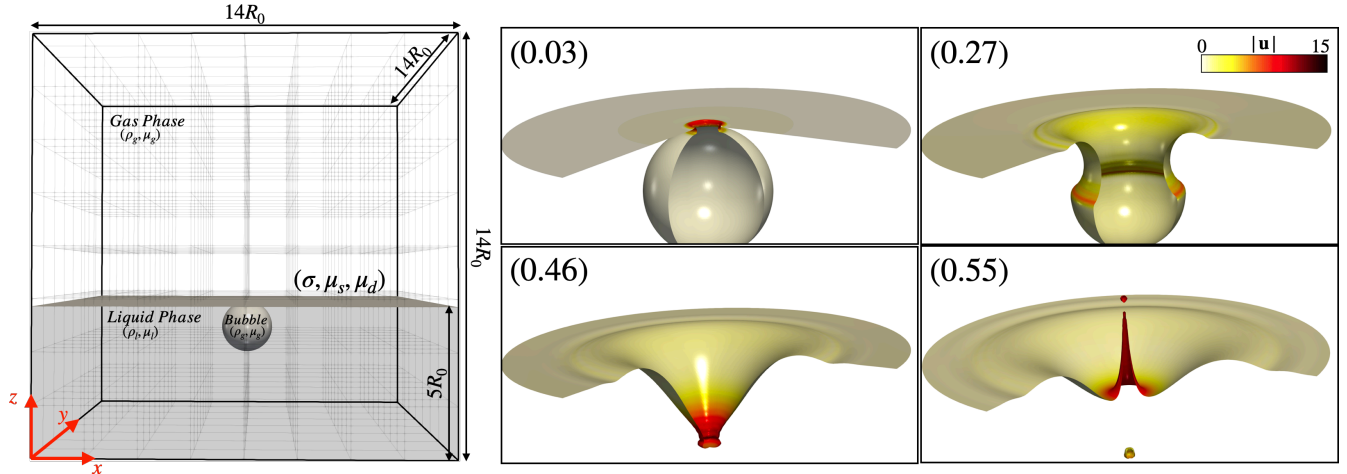


Figure 1: Computational Domain of size $(14R_0)^3$ highlighting the initial shape of the bubble resting close to the interface. The snapshots on the right-hand side outline the 3D evolution of the bubble for $\mu_d = 0$.

where $\tilde{\mathbf{f}}(\tilde{\mathbf{x}}_f)$:

$$\begin{aligned} \tilde{\mathbf{f}}(\tilde{\mathbf{x}}_f) = \int_{\tilde{A}} \tilde{\nabla}_s \cdot [\mathbf{I}_s + (Oh_d - Oh_s)(\tilde{\nabla}_s \cdot \tilde{\mathbf{u}})\mathbf{I}_s \\ + Oh_s(\tilde{\nabla}_s \tilde{\mathbf{u}} \cdot \mathbf{I}_s + \mathbf{I}_s \cdot (\tilde{\nabla}_s \tilde{\mathbf{u}})^T)] \delta(\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_f) d\tilde{A} \end{aligned} \quad (9)$$

Oh_s and Oh_d denote the shear and dilatational Ohnesorge numbers, which are the ratio of surface viscous forces to inertio-capillary forces and Bo is the Bond number, which is the ratio of gravity to surface tension forces. These dimensionless numbers can be calculated by the following equations:

$$Oh = \frac{\mu}{\sqrt{\rho R_0 \sigma}} \quad (10) \quad Oh_d = \frac{\mu_d}{\sqrt{\rho R_0^3 \sigma}} \quad (12)$$

$$Bo = \frac{\rho g R_0^2}{\sigma} \quad (11) \quad Oh_s = \frac{\mu_s}{\sqrt{\rho R_0^3 \sigma}} \quad (13)$$

2.4. Problem setup and numerical method

A sufficiently large cubic computational domain (open from the top, laterally closed (free-slip) and no-flow at the bottom (no-slip)), of $14R_0^3$ as shown in Figure 1 is modelled. In conjunction with previous studies, this choice of domain size ensures that boundary effects and reflected wave propagation do not interfere with the bubble bursting mechanics [7] [8]. We employed a structured uniform grid of resolution 512^3 . We set a uniform grid cell size, $\Delta x = 7 \times 10^{-3}/512 = 1.37 \times 10^{-5} \text{ m}$ ($\frac{R}{\Delta x} = 36.6$) to capture the key indicators in the bursting bubble process. Jet ejection and bubble bursting involve sharp, localised curvature changes

and rapid dynamics, hence the implementation of a highly resolved domain is required [16]. An explicit time-stepping is implemented with a temporal resolution of $\Delta t \sim \mathcal{O}(10^{-8} - 10^{-7})$ s to reduce the discretisation errors in computing the surface gradient of velocities on the interface. A parallelised multigrid approach is implemented for a fast pressure solver and the advection term is solved via essentially non-oscillatory (ENO) scheme. The discretisation and implementation of the surface divergence operation of the Boussinesq-Scriven surface tensor leverages the hybrid level-set based Front tracking method. For more information, readers can refer to [17].

Our controlling parameters in this study are the Bo, Oh, Oh_d and Oh_s . We fixed $Bo = 0.1$ and $Oh = 0.005$, matching the regime investigated by Walls et al. [18] in their study of jet drops from bursting bubbles. At these values, effects of gravity and bulk viscosity on jet formation and droplet production can be considered negligible, isolating the interfacial rheology. Our chosen parameters imitate a 0.5 mm radius bubble at an interface which is fully saturated with surfactant (when the concentration of surfactant in the bulk is equal to or above the critical micelle concentration), i.e. the surface tension, $\sigma = 0.025 \text{ N/m}$. This follows from the research of Sane et al. in their paper on surface tension of flowing soap films [19], where they noted that, at sufficiently high surfactant coverage, the interfacial concentration approaches its saturation value and the equilibrium surface tension approaches an asymptotic value of $\sim 30 \text{ mN/m}$.

Singh et al. [20], in the study of the impact

of surface viscosity on droplet stability, outline the existence of specific surfactants like DPPC (Dipalmitoylphosphatidylcholine) and DMPE (Dimyristoylphosphatidylethanolamine) (lipid monolayers) for which the dilatational surface viscosity is 10^5 times greater than the shear surface viscosity ($Oh_d \gg Oh_s$). The use of such a surfactant is assumed, rendering the $Oh_s \approx 0$.

Table 3 summarises all the constant input values:

Table 3: Constant Input Parameters.

Constants	Value	Unit
R_0	5×10^{-4}	m
g	10	m/s^2
ρ_l	1000	kg/m^3
μ_l	5.59×10^{-4}	Pa s
σ	2.5×10^{-2}	N/m
Bo	0.1	—
Oh	5×10^{-3}	—

In our simulations, we considered four cases, spanning a range of dilatational surface viscosities. This was done by manipulating the Oh_d as outlined in Section 2.3 allowing us to effectively observe how this influences jet evolution. To summarise, assuming an interface fully packed with lipid monolayers, we have constant surface tension (no surface elasticity) and negligible surface shear viscosity, allowing us to isolate dilatational surface viscous effects.

The chosen μ_d and Oh_d values for each case are shown in Table 4. Our results were obtained by performing direct numerical simulations (DNS) on a high-performance computer using code BLUE [1].

Table 4: Dilatational surface viscosities for each case.

Case	$\mu_d[\text{kg/s}]$	Oh_d
0	0.00	0.00
1	2.795×10^{-8}	0.0005
2	2.795×10^{-7}	0.005
3	2.795×10^{-6}	0.05

3. Results

3.1. Performance Indicators

In order to characterise how surface dilatational viscosity influences the bubble bursting dynamics, we defined a set of performance indicators. First, we recorded the maximum height of the jet that formed ($\tilde{H}_j$) and whether it pinched off into droplets as this is the most direct measure of surface viscous effects. We also measured the maximum jet velocity ($\tilde{v}_j$) and the interfacial kinetic energy ($\tilde{E}_k$) in order to quantify the intensity of the jet that forms. Alongside the kinetic energy, we tracked the evolution of the surface energy ($\tilde{E}_s$) and dissipated energy ($\tilde{E}_d$) over time to determine how much energy is lost to the formation of a new surface as a daughter bubble is entrained. Finally, we tracked how the bubble shape varied over time for different viscosities by monitoring the capillary angle (θ_c), the bubble depth ($\Delta\tilde{z}$) and the final crater height ($\tilde{H}_f$). Together these indicators provide a coherent framework for assessing the impact of dilatational viscosity on the dynamics of bubble bursting.

3.2. Role of Surface Viscosity

We began the investigation by first examining bubble bursting at a clean interface (Case 0, $\mu_d = 0$) to establish a reference for the surface dilatational effects. In this baseline case, the dynamics were governed solely by bulk properties without any contribution from surface viscous effects.

Figure 2 outlines the 2D evolution of the cavity collapse and jet evolution for all four cases. After the nucleation of a hole at the liquid interface and subsequent cap rupture [$\tilde{t} = 0.030$], the upper interface rapidly retracts and drives a capillary-inertial collapse of the cavity [$\tilde{t} = 0.268$]. These interfacial retractions act inwards and downwards, eventually focusing to form a crater where the jet velocity reaches its maximum [$\tilde{t} = 0.457$]. Following this, the interfacial curvature reverses, driving fluid upwards and birthing a thin Worthington jet. As this vertical jet begins to evolve, some of the system's kinetic energy is redirected to the formation of a daughter bubble beneath the cavity. Under its own inertial forces, the jet begins to elongate and thin. Once inertial stretching has sufficiently thinned the jet, capillary forces begin to

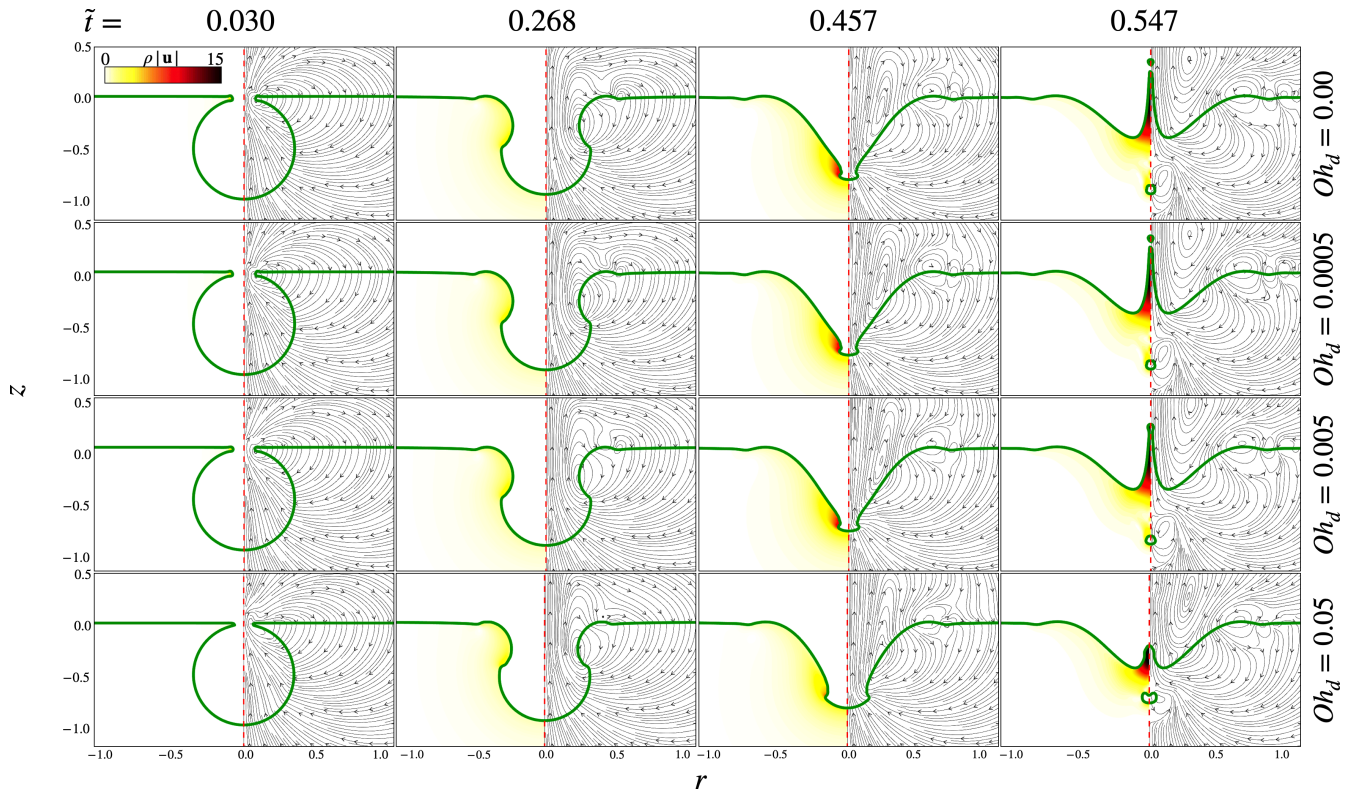


Figure 2: DNS results for Case 0: $Oh_d = 0.00$, Case 1: $Oh_d = 0.0005$, Case 2: $Oh_d = 0.005$, Case 3: $Oh_d = 0.05$

dominate, and surface tension acts to minimise the surface area (Rayleigh-Plateau instability), resulting in pinch-off and the subsequent formation of a droplet [$\tilde{t} = 0.547$].

Qualitatively, bubble bursting dynamics in the clean case ($Oh_d = 0$) and case 1 ($Oh_d = 0.0005$) are indistinguishable. From Figure 2 it is apparent that for both cases, the time of cavity collapse, the time of pinch, and the shape of the bubble over its bursting sequence are all identical. Viscous forces in case 1 are simply too weak to have considerable effects on the bursting bubble mechanics.

Increasing the dilatational viscosity further in Case 2 ($Oh_d = 0.005$) exhibits minute deviations from the clean case. Distinctions between the cases emerge only as the collapsing cavity approaches crater formation. In Case 2, signs of dilatational viscous effects become apparent as the cavity neck widens slightly from $0.356R_0$ in the clean case to $0.412R_0$ [$\tilde{t} = 0.457$]. The increase in dilatational viscosity suppresses the radial inflow and increases the resistance to retraction along the interface. This retards the evolution of the cavity, giving rise to the widening observed. Evidence of this slight delay is accentuated as at pinch off, the jet for $Oh_d = 0.005$ lags behind the clean case [$\tilde{t} = 0.547$]. These

observations indicate that, although there is a visible impact on the bursting sequence, in this case, the dilatational viscosity still remains too small to induce substantial changes in the bursting dynamics.

In Case 3 ($Oh_d = 0.05$) numerous differences in the temporal evolution and intensity of bubble bursting dynamics can be seen. The damping of capillary waves is amplified as seen by the substantial delay in retraction [$\tilde{t} = 0.268$]. The viscosity induced delay, first observed in Case 2, is magnified as the cavity neck widens to $0.862R_0$; over twice as wide as all previous cases [$\tilde{t} = 0.457$]. Although the crater formed at similar times for Cases 0 to 2, in Case 3, much more time is required to overcome the crater and thus form an upward jet. Therefore, it is apparent that at a higher surface dilatational viscosity, surface viscous effects dominate inertial effects. As a result, although other cases approach pinch off [$\tilde{t} = 0.547$], for Case 3 we observe a prematurely developed jet profile.

Having established that increasing dilatational viscosity progressively delays cavity collapse and jet emergence, the following analysis quantifies its impact on the overall jet response.

3.3. Impact of Surface Viscosity on the Jet Response

3.3.1. Energy Profiles

Kinetic energy and dissipated energy profiles provide insights into dilatational viscous behaviours. Figures 3a and 3b outline the evolution of kinetic energy ($\tilde{E}_k$) and dissipated energy ($\tilde{E}_d$) over time for all cases. From the figures, two critical stages in the dynamics can be observed: (1) the crater formation and (2) the subsequent formation of an entrained daughter bubble. In Figure 3a we can see that the kinetic energy reaches its global maximum at $\tilde{t} = 0.457$, corresponding to the time of crater nadir formation. This is rationalised by examining the interfacial dynamics at this instant. During cavity collapse, the capillary waves generated at the rim focus at the cavity bottom, forming a large capillary pressure gradient that generates axial liquid acceleration. This results in a high-velocity region forming around the axis, and the crater nadir forms at the instant when all the downward capillary forces converge. Following the global maximum, all cases show a steep decline in kinetic energy. As the entrained daughter bubble forms below the cavity, kinetic energy is converted to surface energy. In their research on secondary bubble entrainment [11], Ji et al. remark that during cavity-neck thinning and daughter bubble formation, work must be done against surface tension, resulting in some kinetic energy being converted into surface energy. Simultaneously, kinetic energy is dissipated as changes in surface area strongly opposes interfacial rigidity. We conduct the following energy budget:

$$E_i = E_k(t = 0) + E_s(t = 0) \quad (14)$$

$$E_i = E_k(t) + E_s(t) + E_d(t) \quad (15)$$

$$1 = \tilde{E}_k + \tilde{E}_s + \tilde{E}_d \quad (16)$$

where E_i denotes the initial energy, $E_k(t)$ (kinetic energy at each time step) was obtained from the output of the DNS, $E_s(t)$ was calculated as $E_s(t) = \sigma A(t)$ and $A(t)$ is the interfacial surface area also obtained from the DNS output.

From the energy budget, we noted that although the surface energy rose between $\tilde{t} = 0.450$ and $\tilde{t} = 0.480$, the energy dissipated for the same time period was much larger. In Newtonian fluids, the

dissipation rate is defined as $D = \mu \left(\frac{\partial u}{\partial y} \right)^2$ where $\left(\frac{\partial u}{\partial y} \right)$ is the velocity gradient. Therefore, the observed spike in dissipated energy can be attributed to the fact that viscous stresses oppose surface tension-driven deformation.

Differences in energy profiles between cases 0, 1 and 2 are negligible, as depicted in 3a and 3b. Therefore, the discussion turns to differences between these cases collectively, and case 3 ($Oh_d = 0.05$). Case 3 shows persistently lower $\tilde{E}_k$ values relative to the other cases, also rising to its global maximum at a much more gradual rate. The damping, induced by higher surface viscosity, reduces the jet velocity and weakens wave focusing [21], resulting in the observed trend.

As viscous dissipation is cumulative, the dissipated energy increases monotonically for all cases, as observed in Figure 3b. Initially, the rate of increase in dissipated energy is greater in case 3 than in the cases with lower viscosities, attaining a higher final value. However, post crater formation, energy conversion occurs 40% slower for case 3 ($\tilde{t} \sim 0.48$). Amplified dilatational surface viscosity, pacifies interfacial motions thus reducing the rate of energy conversion.

3.3.2. Maximum Jet Velocity, Crater Height and Jet Height

Figures 4a, 4b and 4c present the variation of maximum jet velocity ($\tilde{v}_j$), crater height ($\tilde{H}_f$) and maximum jet height ($\tilde{H}_j$) across Oh_d regimes. For all three metrics, we can see that a clear transition occurs at $Oh = Oh_d$. In the regime where $Oh_d < Oh$, bulk viscous properties dominate and hence minor deviations are observed across all metrics. Conversely, when $Oh_d > Oh$, surface viscosity begins to dominate, hence interfacial area changes and jet formation are highly retarded, resulting in steep decreases in both maximum jet velocity and crater height.

In Figure 4c, analysing the effect of Oh_d on $\tilde{H}_j$, we can see that the trend is reversed when the threshold $Oh = Oh_d$ is crossed. As depicted in Figure 4c's inset, the maximum jet height is defined as the height achieved at the instant just prior to pinch-off. In cases 0-2, this pinch-off occurs at $[\tilde{t} = 0.547]$, but in case 3 the formation of a jet is substantially delayed with pinch-off only observed at $[\tilde{t} = 0.794]$. The 45%

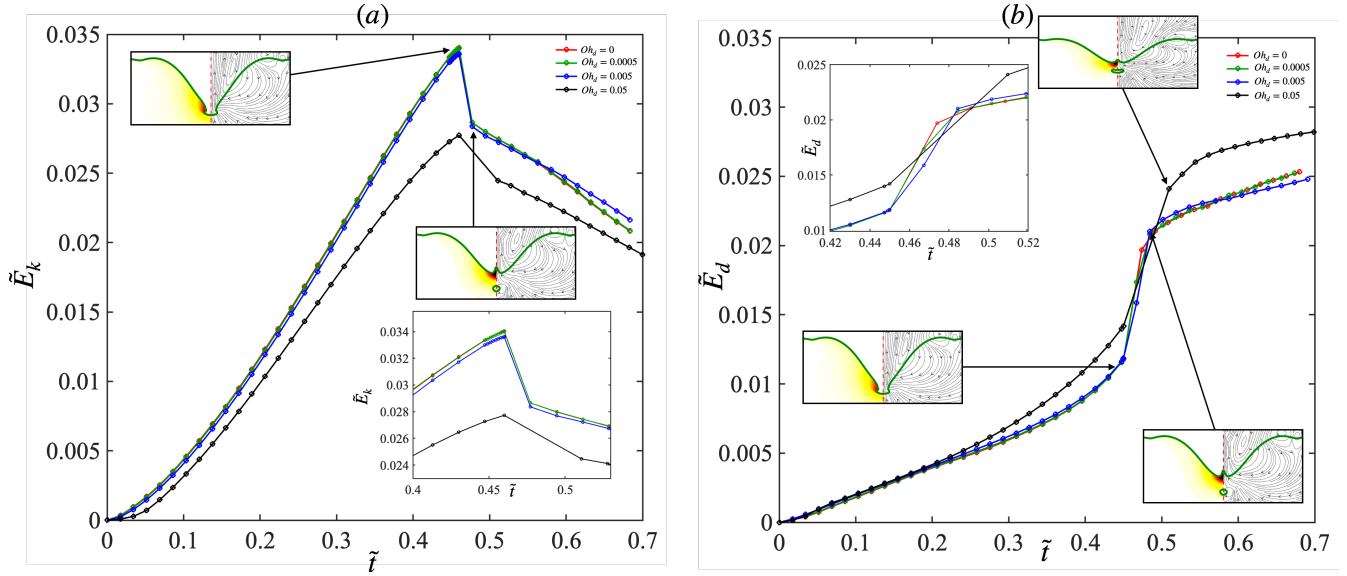


Figure 3: (a) Variation in Interfacial Kinetic Energy over Time (b) Variation in Energy Dissipated over Time

increase in time to pinch-off for case 3 enables more liquid from the bath to enter the jet and subsequently results in a larger height being attained at pinch off.

3.3.3. Bubble Depth and Capillary Angle

In Figures 5a and 5b the temporal evolution of bubble depth ($\Delta\tilde{z}$) and capillary angle (θ_c) are presented. Both these parameters are computed from the time of cap rupture until the instant of crater formation as they characterise the evolution of the cavity. Concurrent with other metrics, differences in cases 0-2 are negligible. In case 3, dilatational surface effects are pronounced, marked by a more gradual reduction, shown by the black curves in Figure 5. This is indicative of the viscous damping that occurs.

4. Discussion

To summarise our results, all metrics exhibit a pronounced change in behaviour when Oh_d is increased to 0.05. Figure 6 shows clearly that at pinch off, we observe a taller, thicker jet that ejects a larger droplet at high dilatational surface viscosity. For this case, we also observe larger energy dissipation and a noticeable delay in jet evolution. To rationalise this behaviour, we examine the tangential velocity and tangential stress (τ) in the plane of the interface, as well as the vorticity field (ω), defined by:

$$\mathbf{u}_t = (\mathbf{u} \cdot \mathbf{t}) \mathbf{t}, \quad \tau = \mu_d [\nabla \mathbf{u} + (\nabla \mathbf{u})^T], \quad \omega = \nabla \times \mathbf{u}. \quad (17)$$

The tangential velocity (u_t) along the interface shows how fast the surface is being stretched or compressed in the direction of capillary-wave propagation. In Figure 7, u_t is plotted as a function of the arc length at two stages of the cavity evolution: (a) during cavity collapse and (b) the instant before the crater nadir forms. In both tangential velocity plots, the circles correspond to the point at which $u_t = 0$, underlining the presence of a stagnation point. These also correspond to the points of maximum curvature as seen in the plot of interfacial height over the radius. The presence of a stagnation point is further substantiated by the tangential stress profiles as at the dotted lines (denoting stagnation points), $\tau = 0$. At a stagnation point, the interface is rigidified and work must be done against this to allow capillary waves to propagate.

From Equation 17, we can see that τ is directly proportional to the dilatational viscosity, so increasing μ_d increases the viscous traction along the interface. In the tangential stress profile, compared to other cases which remain near flat, the black curve (case 3) in Figure 7 oscillates significantly about the stagnation point. The resulting magnitude difference in stress for the highly viscous case gives rise to a considerably more rigid interface at the stagnation point. As a result, there is greater resistance to any further stretching and compression of the interface.

The behaviour of the opposing motions acting at the interface is accentuated in the vorticity colour map shown in Figure 7c. The figure compares the vorticity profile at the instant prior to pinch-off for $Oh_d = 0.00$

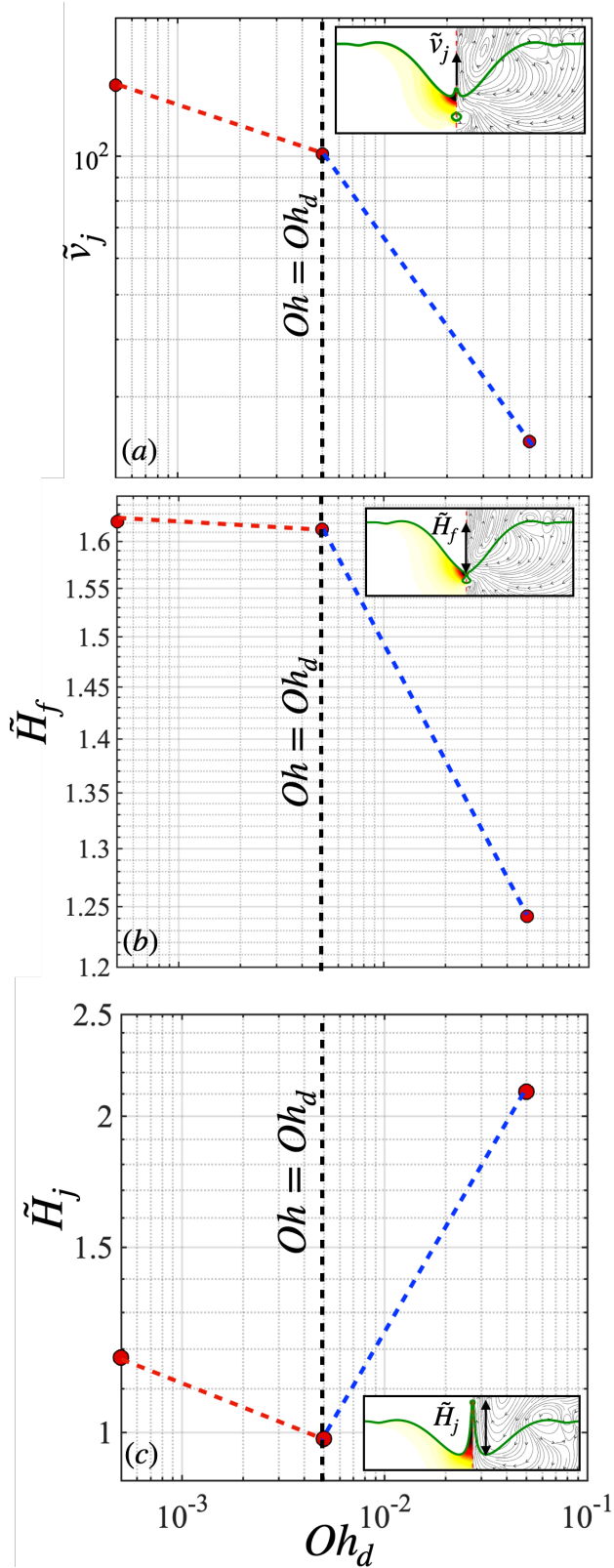


Figure 4: Plots showing the Role of Oh_d on Jet Evolution: (a) Variation in Maximum Jet Velocity (b) Variation in Crater Height (c) Variation in Maximum Jet Height

and $Oh_d = 0.05$. In the figure, a thin neutral vorticity band (light white region) marks the stagnation point.

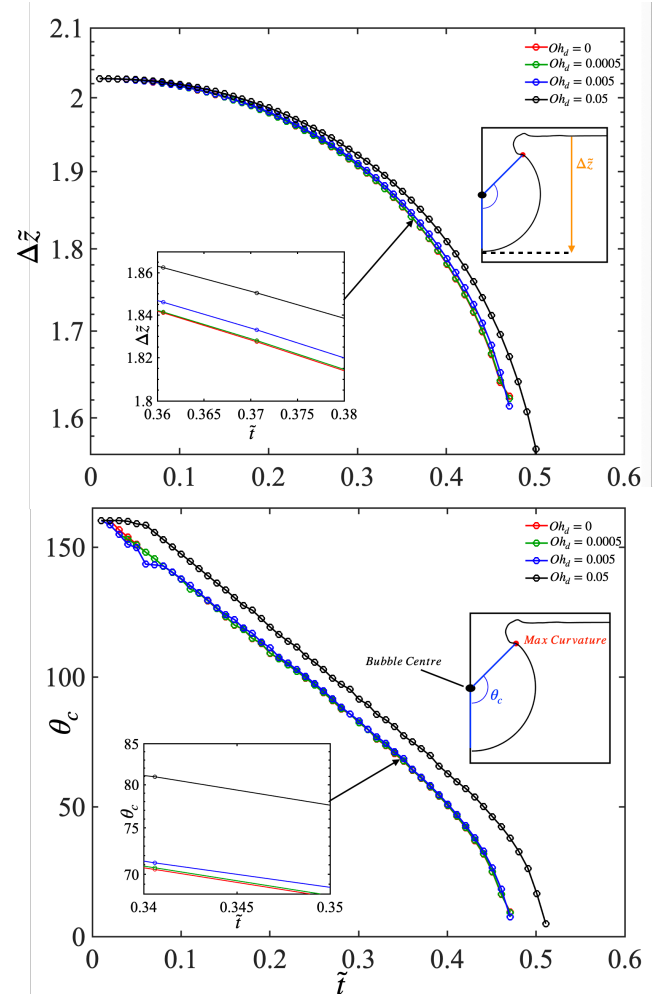


Figure 5: Plots showing the Role of Oh_d on Cavity Evolution: (a) Top Figure showing Variation in Bubble Depth (b) Bottom Figure showing Variation in Capillary Angle

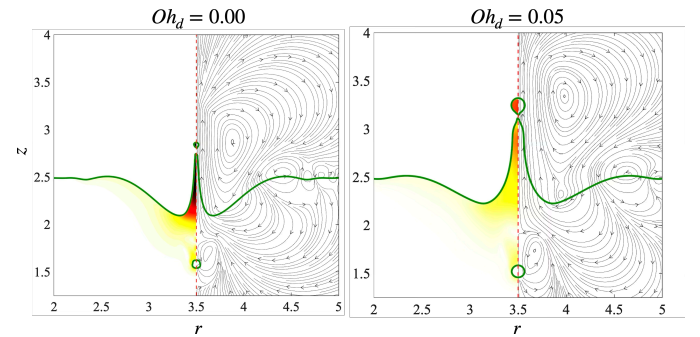


Figure 6: Comparison of Case 0 and Case 3 at Pinch Off

Immediately above and below this band, the vorticity takes large positive and negative values. For $Oh_d = 0.05$, these lobes occupy a larger region and showcase greater magnitudes than in the clean case, which indicates that increased viscous stress give rise to stronger rotational motion.

This increased rigidification drives the stark differences in the behaviour of case 3 as documented

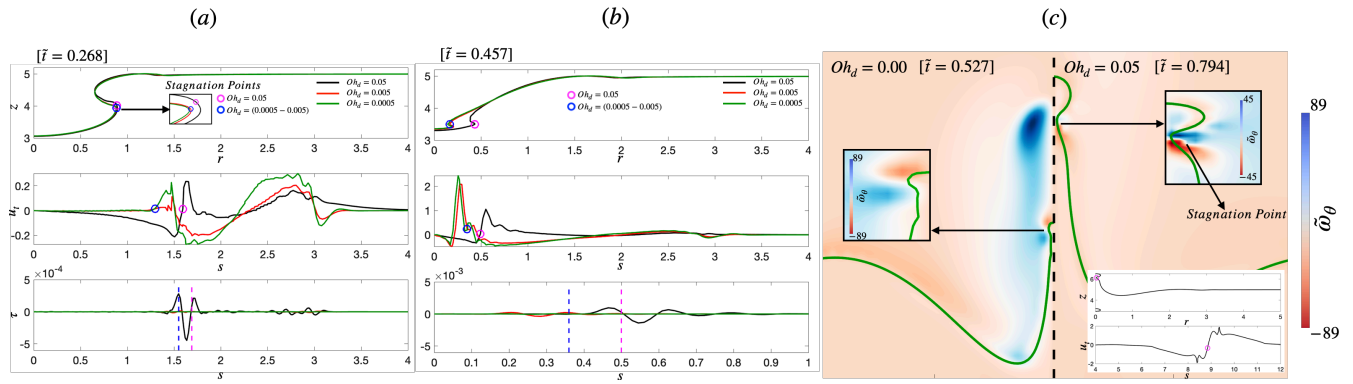


Figure 7: (a) Tangential Velocity and Tangential Stress Profiles along Arc Length during Capillary Wave Propagation (b) Tangential Velocity and Tangential Stress Profiles along Arc Length at the Crater Nadir (c) Vorticity Plot comparing Case 0 and Case 3 at the instant prior to pinch-off

in our findings in Section 3. Elevated resistance, due to a more rigid interface, requires more work to drive capillary wave focussing at the axis. This results in two key outcomes: the first is that more energy is dissipated as the cavity evolves. The second is the existence of a significant delay in evolution without drastically varying the bubble bursting mechanism itself. This results in a more gradual growth and over time, more liquid from the bath accumulates in the Worthington jet. This increases the jet's height and width, consequently ejecting a larger droplet of radius $68.5\mu\text{m}$ (which is nearly three times greater than the clean case, $20.6\mu\text{m}$).

To the best of our knowledge the large positive vorticity observed well above the jet for $Oh_d = 0.00$ in Figure 7c does not arise due to any viscous effects. This rather can be attributed to the inner-gas dynamics of bursting bubbles. In their paper investigating this phenomenon, Singh et al. [22] outline that the rapid expulsion of a bubble gas during cavity collapse generates a high-speed gas jet above the free surface. They observed that the associated gas-gas and gas-liquid shear layers produce intense vorticity in the gas region above the emerging Worthington jet. We believe that similar mechanisms give rise to the observed strong vorticity in our simulation.

5. Conclusion

This work presents the first direct numerical simulations that capture the effects of dilatational surface viscosity on bubble bursting in 3D. As Oh_d exceeds Oh , striking differences in jet evolution and interfacial deformation are observed. Our

results reveal that, at considerable surface dilatational viscosities, there is increased interfacial stiffness and strong resistance to changes in local surface area. This is in accordance with the findings of Ubal et al. in 2005 [23]. Consequently, slower retraction rates, taller Worthington jets and larger ejected droplets are observed as distinct characteristics of viscous dominated interfaces. Deike et al. similarly observed amplified jet drop sizes in the presence of insoluble surfactant during their 2025 study [24]. Potential future studies could further investigate the regime where $Oh_d > Oh$ to determine if there is a threshold beyond which droplet ejection is inhibited. It should be noted that resolutions higher than (512^3) should be employed, as the jagged profile observed in Figure 7 occurred due to insufficient resolution [$s \sim 2.8$]. Computational cost is also a key consideration; low viscous cases were completed in 200 hours and higher viscous cases required approximately 400 hours. Identification of a regime in which droplet ejection is inhibited could enable the deployment of lipid monolayers and other similar surfactants in toxic waste waters, where aerosol release poses serious health risks.

6. Acknowledgements

The authors would like to express their gratitude to the Matar Fluids Group for providing the opportunity to conduct this research. Special thanks go to Dr. Lyes Kahouadji and PhD Debashis Panda for their valuable technical assistance and insights.

References

- [1] S. Shin, J. Chergui and D. Jurić, 'A solver for massively parallel direct numerical simulation of three-dimensional multiphase flows,' *Journal of Mechanical Science and Technology*, vol. 31, pp. 1739–1751, Apr. 2017. DOI: [10.1007/s12206-017-0322-y](https://doi.org/10.1007/s12206-017-0322-y)
- [2] A. H. Woodcock, C. F. Kientzler, A. B. Arons and D. C. Blanchard, 'Giant condensation nuclei from bursting bubbles,' *Nature*, vol. 172, pp. 1144–1145, 1953.
- [3] L. Duchemin, S. Popinet, C. Josserand and S. Zaleski, 'Jet formation in bubbles bursting at a free surface,' *Physics of Fluids*, vol. 14, no. 9, pp. 3000–3008, 2002. DOI: [10.1063/1.1494072](https://doi.org/10.1063/1.1494072)
- [4] A. Moallemi, A. Alberello and I. Thurnherr, 'Links between atmospheric aerosols and sea state in the arctic ocean,' *Atmospheric Environment*, vol. 338, p. 120 844, 2024. DOI: [10.1016/j.atmosenv.2024.120844](https://doi.org/10.1016/j.atmosenv.2024.120844). [Online]. Available: <https://infoscience.epfl.ch/entities/publication/0e3737b1-d616-4309-beea-6e89ee895e44>
- [5] J. M. Boulton-Stone and J. R. Blake, 'Gas bubbles bursting at a free surface,' *Journal of Fluid Mechanics*, vol. 254, pp. 437–466, 1993. DOI: [10.1017/S0022112093002130](https://doi.org/10.1017/S0022112093002130)
- [6] V. Sanjay, D. Lohse and M. Jalaal, 'Bursting bubble in a viscoplastic medium,' *Journal of Fluid Mechanics*, vol. 922, A2, 2021. DOI: [10.1017/jfm.2021.489](https://doi.org/10.1017/jfm.2021.489). [Online]. Available: <https://arxiv.org/abs/2101.07744>
- [7] A. Pico et al., 'Surfactant-laden bubble bursting: Dynamics of capillary waves and worthington jet,' *Physical Review Fluids*, vol. 9, no. 8, p. 083 606, 2024. DOI: [10.1103/PhysRevFluids.9.083606](https://doi.org/10.1103/PhysRevFluids.9.083606). [Online]. Available: <https://journals.aps.org/prfluids/abstract/10.1103/PhysRevFluids.9.083606>
- [8] C. R. Constante-Amores et al., 'Dynamics of a surfactant-laden bubble bursting through an interface,' *Journal of Fluid Mechanics*, vol. 911, A57, 2021. DOI: [10.1017/jfm.2021.164](https://doi.org/10.1017/jfm.2021.164). [Online]. Available: <https://www.cambridge.org/core/journals/journal-of-fluid-mechanics/article/dynamics-of-a-surfactantladen-bubble-bursting-through-an-interface/E3AE192F73A44A60853C7948D695FB5C>
- [9] E. Ghabache and T. Séon, 'Size of the top jet drop produced by bubble bursting,' *Physical Review Fluids*, vol. 1, no. 5, p. 051 901, 2016. DOI: [10.1103/PhysRevFluids.1.051901](https://doi.org/10.1103/PhysRevFluids.1.051901). [Online]. Available: <https://doi.org/10.1103/PhysRevFluids.1.051901>
- [10] Z. Yang, B. Ji and J. Feng, 'Effect of surface viscoelasticity on top jet drops produced by bursting bubbles,' *Soft Matter*, 2024. DOI: [10.1039/d4sm00243a](https://doi.org/10.1039/d4sm00243a). [Online]. Available: <https://pubs.rsc.org/en/content/articlelanding/2024/sm/d4sm00243a>
- [11] W. Ji, S. Talukdar, M. Trejo, O. A. Basaran and K. Pan, 'Secondary bubble entrainment via primary bubble bursting at a viscoelastic surface,' *Physical Review Letters*, vol. 131, no. 10, p. 104 002, 2023. DOI: [10.1103/PhysRevLett.131.104002](https://doi.org/10.1103/PhysRevLett.131.104002)
- [12] E. J. Vega et al., 'Influence of a surfactant on bubble bursting,' *Chemical Engineering Science*, vol. 151, no. 111097, pp. xxx–xxx, 2024. DOI: [10.1016/j.2024.cesxxx](https://doi.org/10.1016/j.2024.cesxxx)
- [13] D. Dey, J. M. Boulton-Stone, A. N. Emery and J. R. Blake, 'Experimental comparisons with a numerical model of surfactant effects on the burst of a single bubble,' *Chemical Engineering Science*, vol. 52, no. 24, pp. 4785–4795, 1997. DOI: [10.1016/S0009-2509\(97\)00083-3](https://doi.org/10.1016/S0009-2509(97)00083-3)
- [14] J. M. Boulton-Stone and J. R. Blake, 'The effect of surfactant on bursting gas bubbles,' *Journal of Fluid Mechanics*, vol. 548, pp. 281–313, 1995. DOI: [10.1017/S0022112005007501](https://doi.org/10.1017/S0022112005007501)
- [15] S. Shin, J. Chergui, D. Juric, L. Kahouadji, O. K. Matar and R. V. Craster, 'A hybrid interface tracking–level set technique for multiphase flow with soluble surfactant,' *Journal of Computational Physics*, vol. 359, pp. 409–435, 2018. DOI: [10.1016/j.jcp.2018.01.010](https://doi.org/10.1016/j.jcp.2018.01.010)
- [16] J. Hua, J. F. Stene and P. Lin, 'Numerical simulation of 3d bubbles rising in viscous liquids using a front tracking method,' *Journal of Computational Physics*, vol. 227, no. 6, pp. 3358–3382, 2008. DOI: [10.1016/j.jcp.2007.12.002](https://doi.org/10.1016/j.jcp.2007.12.002). [Online]. Available: https://discovery.dundee.ac.uk/ws/portalfiles/portal/82972/JCP_Paper_manuscript_V7.pdf
- [17] D. Panda et al., 'Direct numerical simulation of two-phase flows with surfactant-induced surface viscous effects,' *arXiv preprint arXiv:2509.24722*, 2025.
- [18] P. L. L. Walls, L. Henaux and J. C. Bird, 'Jet drops from bursting bubbles: How gravity and viscosity couple to inhibit droplet production,' *Physical Review E*, vol. 92, no. 2, p. 021 002, 2015. DOI: [10.1103/PhysRevE.92.021002](https://doi.org/10.1103/PhysRevE.92.021002)
- [19] A. Sane, S. Mandre and I. Kim, 'Surface tension of flowing soap films,' *Journal of Fluid Mechanics*, vol. 841, R2, 2018. DOI: [10.1017/jfm.2018.28](https://doi.org/10.1017/jfm.2018.28)
- [20] N. Singh and V. Narsimhan, 'Impact of surface viscosity on the stability of a droplet translating through a stagnant fluid,' *Journal of Fluid Mechanics*, vol. 927, 2021, Published online 01 October 2021.
- [21] Y. Cheng, Y. Shen, D. Liu, J. Xu and Y. Sui, 'Numerical analysis of bubble bursting at the liquid surface by wave propagation,' *International Journal of Thermal Sciences*, vol. 152, p. 106 341, 2020. DOI: [10.1016/j.ijthermalsci.2020.106341](https://doi.org/10.1016/j.ijthermalsci.2020.106341)
- [22] D. Singh and A. K. Das, 'Dynamics of inner gas during the bursting of a bubble at the free surface,' *Physics of Fluids*, vol. 33, no. 5, p. 052 105, 2021. DOI: [10.1063/5.0048121](https://doi.org/10.1063/5.0048121)
- [23] S. Ubal, M. D. Giavedoni and F. A. Saita, 'Influence of surface viscosity on two-dimensional faraday waves,' *Industrial & Engineering Chemistry Research*, vol. 44, no. 4, pp. 1090–1099, 2005. DOI: [10.1021/ie049188b](https://doi.org/10.1021/ie049188b). [Online]. Available: <https://pubs.acs.org/doi/10.1021/ie049188b>
- [24] J. Eshima, T. Aurégan, P. K. Farsoiyya, S. Popinet, H. A. Stone and L. Deike, 'Size amplification of jet drops due to insoluble surfactants,' *Physical Review Fluids*, 2025, Accepted manuscript. DOI: [10.1103/t9k3-dqyc](https://doi.org/10.1103/t9k3-dqyc). [Online]. Available: <https://doi.org/10.1103/t9k3-dqyc>

Machine Learning Enhanced Geometric Optimisation of Crystallisation Reactors

Marco Barbacci-Schettino & Thomas Slade^a

^a*Department of Chemical Engineering, Imperial College London,*

Abstract

Paracetamol is produced through reactive crystallisation, where reactor geometry influences local hydrodynamics and mixing which in turn determine the particle size distribution (PSD). A narrow, uniform PSD is essential to enhance compressibility properties and tablet quality. Traditional reactor designs rely on low-dimensional, handcrafted geometric parameterisations that cannot depict the complex geometries now available via additive manufacturing. This makes systematic optimisation of crystallisation performance computationally expensive and prohibitive. This study introduces a compute-aware multi-fidelity Bayesian optimisation (MFBO) framework that optimises reactor geometry over a learned Variational Autoencoder (VAE) shape manifold, coupling OpenQBMM crystallisation physics with a robust PSD surrogate to discover geometries that produce uniform PSDs at lower computational costs. A convolutional VAE is trained on 3,000 synthetically generated 2D reactor silhouettes (256×256 pixels), compressing complex cross-sections into a 30-dimensional latent space. This latent space is then initialised with Latin hypercube sampling and explored using the MFBO. Each candidate reactor design is then evaluated using computational fluid dynamics (CFD) and population balance modelling via a coupled OpenFOAM–OpenQBMM model. Simulations are performed under laminar flow with growth-dominated crystallisation kinetics and mesh fidelities varying from 4,000 to 12,250 cells. The resulting PSD is quantitatively assessed using a tanks in series parameter, N , with larger values indicating narrower distributions. This uniformity is maximised within a budget of 50 MFBO evaluations and 18 hours of wall clock time. This framework enhances a nominal reactor design from $N = 11.82$ to an optimal solution of $N = 134.03$. Under these conditions, the learned geometries alter residence time distributions across the reactor to promote a narrow, nearly unimodal crystal size distribution. The proposed open-source methodology is directly applicable to 3D geometries and other crystallisation systems where geometry performance is complex and high-dimensional.

Keywords: Reactor Design, Crystallisation, Multi-fidelity Bayesian optimisation, Variational autoencoder, Computational Fluid Dynamics

1. INTRODUCTION

Paracetamol (Acetaminophen) is one of the most widely used pain-relief medications, with production rates exceeding 275,000 tonnes globally in 2024 [1]. Its synthesis involves the acylation reaction between 4-aminophenol and acetic anhydride in an aqueous medium with acetic acid as a byproduct [2]. The product is isolated through reactive crystallisation. Reactor geometry plays a significant role in determining hydrodynamics, mixing characteristics, and ultimately, the PSD of paracetamol crystals. Non-uniform PSD can lead to tablet defects during the manufacturing process such as capping [3].

Innovations in machine learning (ML) now enable the generation of highly complex reactor geometries that were previously unattainable. This substantially increases the design space available for crystallisation process optimisation. Exploring this design space with CFD simulations integrated with Population Balance Equation (PBE) frameworks like OpenQBMM [4] is computationally expensive, even for a single evaluation. It is therefore physically infeasible to test every possible geometrical configuration. Consequently, data-driven approaches have emerged as increasingly viable alternatives [5]. These methods, however, are intractable at higher dimensions, making effective parameterisation essential for data-driven optimisation.

Conventional geometric optimisation techniques rely on intuitively selected parameters and struggle to represent the geometric complexity, resulting in suboptimal reactor candidates. This restricts the design space and inhibits the identification of optimal reactor designs. These limitations are more pronounced in multidimensional space, where performance parameters often conflict with each other, are inefficient and scale poorly. Previous work has demonstrated the effectiveness of ML-driven process intensification, including the use of Artificial Neural Networks [6] and Gaussian processes [7][8] to model transport performance in reactors, mixers and heat exchangers. Unlike previous studies, [6][7][8] that rely on manual parameterisation. This study proposes a framework capable of learning cross-sectional parameterisation into reduced-dimensional latent representations. Enabling silhouettes of proposed geometries to be accurately parameterised. This allows the identification of reactor geometries that yield paracetamol crystals with optimised PSD, utilising the full potential of modern ML-driven design.

2. Problem Formulation

When dealing with high-dimensional and expensive black box optimisation problems, such as optimisation of crystallisation reactors, efficient parameterisation is needed to gener-

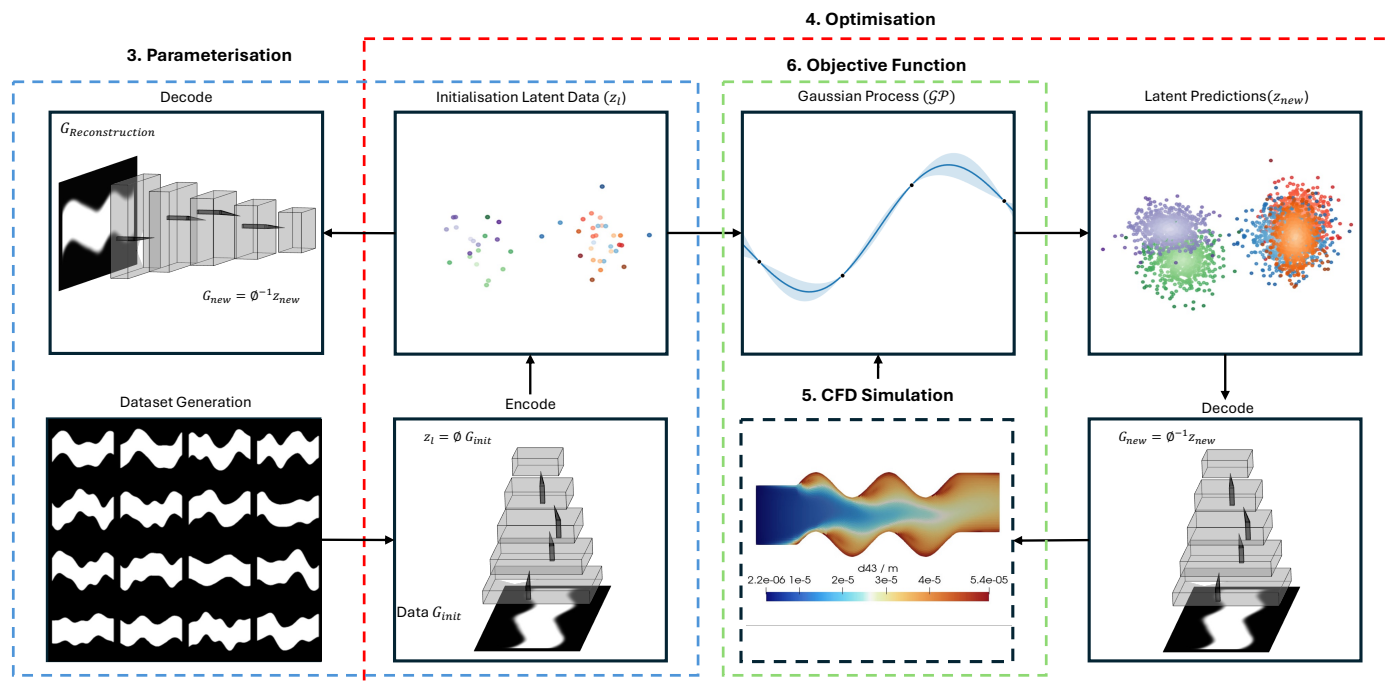


Figure 1: Overview of the optimisation framework, highlighting the key sections of the paper involved in generating, simulating and identifying geometries that maximise PSD uniformity

ate novel and optimised geometries. Parameterisation reduces the dimensionality of complex structures. For example, 3D coiled reactors are commonly parameterised using variables like radius, pitch, length, and width. Numerous studies have successfully used this parameterisation technique. Izadi et al. demonstrated the methodology’s effectiveness for microheli-cally coiled tubes [9], while Savage et al. extended it by including volume constraints as an additional design parameter [8]. Mansour et al. further validated this method over a wider range of parameters [10]. Parameterisation techniques varying radius pitch and length of coiled reactors are particularly useful because they allow optimisation algorithms to operate on coiled geometries without requiring volume intersections or geometric constraints. As these parameters are dimensionless, fundamental relationships governing crystal growth and PSD are maintained during the optimisation process and guarantee to remain robust across different reactor scales.

However, recent parameterisation advances have primarily focused on general shape optimisation rather than detailed cross-sectional geometry. Cross-sectional parameterisation requires higher dimensionality due to the greater complexity. This work outlines a novel parametric technique applied to 2D reactors, which can be naturally extended to 3D reactor geometries.

A convolutional VAE is used to encode geometric silhouettes of a reactor into a reduced-dimensionality latent space, as shown in Figure 1. A VAE fits the manifold over a Gaussian distribution, providing a continuous, reduced dimensionality design space capable of representing highly complex reactor geometries. To the authors’ knowledge, this is the first use of VAEs for reactor parameterisation, whereas previous ap-

plications have focused on molecular design [11]. This is integrated with the reactive crystallisation model within OpenFOAM using the OpenQBMM quadrature-based population balance framework, enabling the prediction of paracetamol crystal growth kinetics. Coupling this solver with a VAE-based latent geometry representation and a cost-aware MFBO algorithm enables systematic and efficient exploration of high-dimensional crystallisation reactor design spaces.

3. Parameterisation

Uniting deep neural networks with variational inference, VAEs have enhanced traditional dimensionality reduction methods to obtain a more meaningful latent space. They form a robust class of models capable of learning unobserved representations in complex datasets. Rather than mapping each input to a single latent point like in a deterministic autoencoder, a VAE predicts the parameters of an approximate posterior conditioned on the input data [12]. This enables virtually any reactor-like shape to be learned and generated from a latent space with a dimensionality of 30, representing a marked improvement over the cross-sectional parameterisation methods employed by Savage et al. [8].

VAEs have been applied to a wide range of domains, from recommender systems [13] to language and sequence modelling [12] and image generation [14]. Within process intensification, they are typically used for chemical design. This is done by mapping discrete chemical language strings (SMILES) into a continuous latent manifold. Property predictors and Bayesian optimisation can then steer searches toward novel, valid molecules with desirable properties [11]. To the authors’

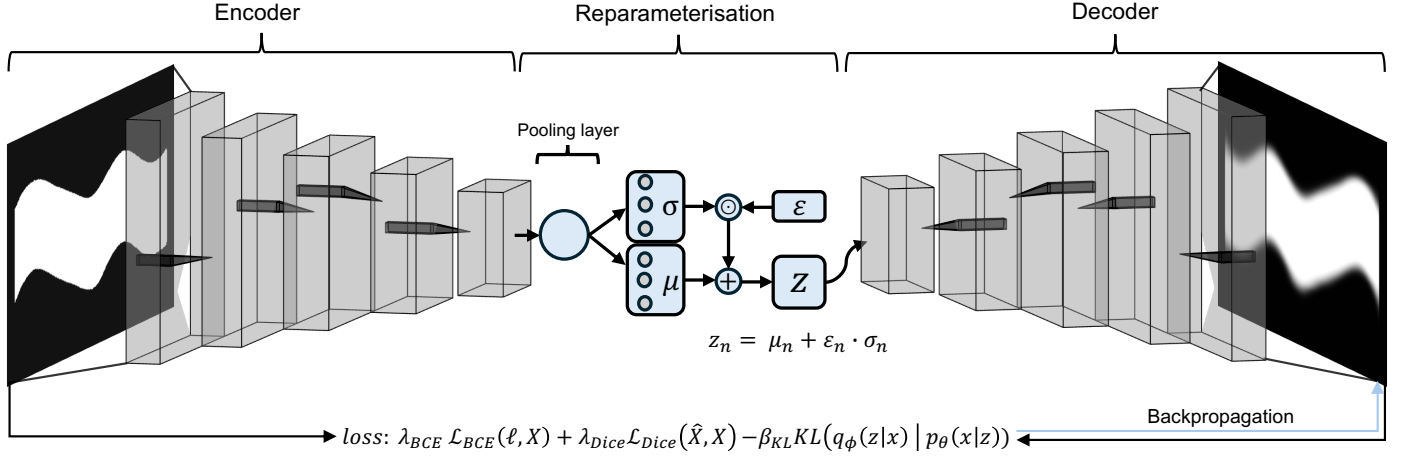


Figure 2: Variational Autoencoder (VAE) architecture used for geometry parameterisation

knowledge, there is no implementation of VAEs in process intensification. Therefore, this study proposes a VAE-based parameterisation framework which utilises a dimensionality reduction whilst maintaining a richer latent space.

3.1. Reactor Design

A synthetic dataset was generated using a base sine profile and a Fourier sum of cosine and sine terms. The sine equation was selected as a nominal geometry for its simplicity, while the Fourier series was used to create asymmetric variations of this baseline. This procedure yielded a synthetic dataset of 3,000 reactor geometries constrained within a 10% surface area tolerance range. As the Fourier series has far more degrees of freedom than the base sine term, it can generate a larger number of distinct geometries, minimising repeating geometries in the dataset. For this reason, 500 geometries were generated from the sine-based profile and 2,500 from the Fourier sum, creating a dataset large enough for effective VAE training with reactor geometries. Example geometries that were generated can be seen in Figure 1.

3.2. Variational Autoencoders

Considering a latent space model with prior $p(\mathbf{z})$ and likelihood $p_\theta(\mathbf{x} | \mathbf{z})$, where θ are decoder parameters and $\mathbf{z} \in \mathbb{R}^d$ are unobserved latent space points. The marginal log-likelihood is intractable for deep autoencoders, as it would involve integrating over the latent space $\mathbf{z}$ [15]. Instead, the posterior is estimated with another distribution $q_\phi(\mathbf{z} | \mathbf{x})$, with parameters ϕ . The distribution is determined such that a similarity measure is minimised; in this case, the Kullback–Leibler divergence (KL) (1) is used.

$$\text{KL}(q(\mathbf{z} | \mathbf{x}) || p(\mathbf{z})) = \int q(\mathbf{z} | \mathbf{x}) \log \frac{q(\mathbf{z} | \mathbf{x})}{p(\mathbf{z})} d\mathbf{z} \quad (1)$$

The choice of distribution reflects the type of data input, where typically a Gaussian distribution is used for real-valued data. The Evidence Lower Bound (ELBO) can now be used as a tractable lower bound on the marginal log likelihood,

$$\log p_\theta(\mathbf{x}) \geq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x} | \mathbf{z})] - \text{KL}(q_\phi(\mathbf{z} | \mathbf{x}) || p(\mathbf{z})) \quad (2)$$

This tractable lower bound now gives an expectation of the likelihood over the posterior distribution and the KL divergence. The KL term acts as a regulariser that pushes the posterior towards the prior, with a smaller KL term allowing for more informative latents. To obtain low-variance gradients through stochastic nodes, the reparameterisation trick [16] [17] expresses latent samples as deterministic transformations of noise:

$$\mathbf{z} = \mu_\phi(\mathbf{x}) + \sigma_\phi(\mathbf{x}) \odot \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (3)$$

where $\mu_\phi, \sigma_\phi : \mathcal{X} \rightarrow \mathbb{R}^d$ are encoder outputs and $\odot$ denotes elementwise multiplication. This is all incorporated in the final loss function for the likelihood over the posterior distribution. The final loss function is given as follows:

$$\mathcal{L}(\theta, \phi; \mathbf{x}) = \frac{1}{L} \sum_{\ell=1}^L \log p_\theta(\mathbf{x} | \mathbf{z}) - \beta_{\text{KL}} \text{KL}(q_\phi(\mathbf{z} | \mathbf{x}) || p(\mathbf{z})) \quad (4)$$

Since the ELBO is a lower bound on the log-likelihood of the input data, a maximisation approach in training provides the VAE with generative capabilities. In order to estimate the posterior, the KL divergence needs to be minimised and therefore, the KL term in 4 is set to be negative. Posterior collapse during training occurs when the approximate posterior becomes nearly identical to the Gaussian prior. This effectively reduces the latent space to a Gaussian distribution and causes the model to ignore information from the input data $\mathbf{x}$. To mitigate this, the β_{KL} term is utilised to reduce regularisation, preventing posterior collapse as initially introduced in [18]. The weighted term can be used to balance regularisation and reconstruction quality, with literature suggesting values of $\beta \leq 0.5$ are preferable.

This surrogate objective has proven effective in generative spaces. In this work, it has been coupled with convolutional neural networks to map binary reactor silhouettes $\mathbf{x} \in \{0, 1\}^{1 \times 256 \times 256}$ into a low-dimensional latent representation

$\mathbf{z} \in \mathbb{R}^{d_z}$ as seen in Figure 2. The encoder maps 256×256 binary reactor images to a 4×4 bottleneck while increasing the number of feature channels. It consists of six successive Down blocks that repeatedly apply two 3×3 convolutions with padding, group normalisation with 8 groups and a SiLU activation, followed by a stride-2 convolution for spatial downsampling. Starting from a single input channel, the channel width is increased through intermediate stages up to $C = 768$, so that the final encoder output is a tensor $h \in \mathbb{R}^{C \times 8 \times 8}$. A global average pooling layer collapses this to a vector $g \in \mathbb{R}^C$, which is then passed through two fully-connected layers to obtain the mean $\mu_\phi(\mathbf{x})$ and log-variance $\log \sigma_\phi^2(\mathbf{x})$ of a diagonal Gaussian posterior in an $d_z = 30$ -dimensional latent space.

The decoder mirrors this structure in reverse, where the latent vector $\mathbf{z} \in \mathbb{R}^{30}$ is mapped through a fully-connected layer to obtain a $C \times 4 \times 4$. Then, six Up blocks are applied that alternate nearest-neighbour upsampling by a factor of two with pairs of 3×3 convolutions, group normalisation and SiLU activations, gradually reducing the channel width back to 8. A final 3×3 convolution produces a single-channel logit field $\ell \in \mathbb{R}^{1 \times 256 \times 256}$. This encoder-decoder construction remains lightweight enough for repeated evaluation inside the Bayesian optimisation loop while accurately reconstructing sharp binary reactor geometries.

In particular, the reconstruction loss is defined as a weighted combination of a binary cross-entropy (BCE) term and a Dice loss as seen in Figure 2. Here ℓ are the decoder logits, which are transformed via the sigmoid function, yielding pixel-wise probabilities $\hat{\mathbf{x}} = \text{Sig}(\ell)$. The BCE term is implemented directly on the logits for numerical stability in high-resolution binary images.

$$\mathcal{L}_{\text{BCE}}(\ell, \mathbf{x}) = -\frac{1}{N} \sum_{i=1}^N [x_i \log \text{Sig}(\ell_i) + (1 - x_i) \log(1 - \text{Sig}(\ell_i))] \quad (5)$$

where N is the number of pixels. To better handle the strong class imbalance between the reactor region and the background, a Dice loss is utilised. Originally developed for medical image segmentation and analysis [19], the Dice loss improves the boundary reconstruction.

$$\mathcal{L}_{\text{Dice}}(\hat{\mathbf{x}}, \mathbf{x}) = 1 - \frac{2 \sum_{i=1}^N x_i \hat{x}_i + \varepsilon}{\sum_{i=1}^N x_i + \sum_{i=1}^N \hat{x}_i + \varepsilon} \quad (6)$$

where ε is a small constant for numerical stability. The BCE term enforces pixel-wise fidelity, and the relative weights λ_{BCE} and λ_{Dice} are chosen to balance these two effects.

Figure 3 shows the effects of each term in the loss function. During model validation, a test set of 250 reactor geometries was used to avoid validation set bias during hyperparameter optimisation. Three geometries are sampled from the test set and visualised in Figure 3.a. Figures 3.b and 3.c show reconstructions from a BCE loss function and the proposed loss function with Dice loss. It was found that the Dice loss term improved both the general shape reconstruction and reduced the standard deviation of Gaussian noise at the reactor boundaries. In order to maintain a continuously differentiable loss function, the images are not treated as binary images, but instead as pixel-wise

probabilities. This can be intuitively interpreted as grey-scaled images, and for silhouette-like geometries, Gaussian noise is most prevalent at the boundaries. This raises issues in post-processing for meshing, and therefore, a threshold of 0.5 is applied to the probabilities to further remove the noise, as seen in Figure 3.d.

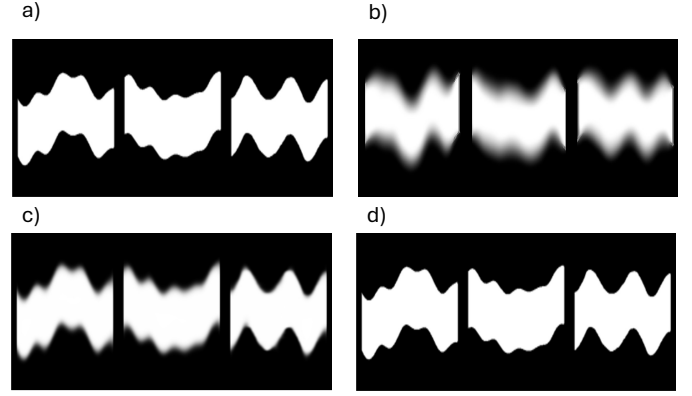


Figure 3: Visual example of the Dice loss: a) shows the VAE input, b) shows the reconstructed images using BCE and KL losses, c) shows the reconstructed images using a BCE, Dice and KL loss, d) shows the post-processing of figure c) when adding a threshold of 0.5.

Hyperparameter optimisation was performed with Hyperopt and validated on the test set. VAE training was carried out on 32 CPUs, a single RTX6000 GPU to aid with training, and 80 GB of RAM for 200 epochs with a batch size of 64, a learning rate of $3e^{-3}$ and a latent space dimensionality of 30. A weight of 1 was set for the BCE loss and 0.5 for the Dice loss, whilst the KL loss was scheduled using a sigmoid function over the iterations with a maximum weighting of 0.3. This setup was chosen to stabilise training and promote reconstruction validity.

This weighted BCE, Dice and KL objective has proven effective for learning robust generative latent spaces for binary shape data, particularly when coupled with convolutional architectures for image modelling.

4. Optimisation

When approaching an optimisation problem with a physics-based model such as CFD, computational expense must be carefully considered. Model complexity and cost can be controlled by varying the simulation fidelity. Using a larger number of cells to evaluate flow fields and PBEs increases accuracy but also substantially increases computational expense. High-dimensional Bayesian optimisation typically requires a greater number of evaluations to converge to an optimal solution due to the complexity of the search space. MFBO mitigates this by using lower fidelity simulations to explore the search space and higher fidelities when exploiting promising regions, making a previously intractable problem viable under computational budget constraints.

The MFBO routine is implemented in Python using BoTorch and GPyTorch. The core loop operates directly in the latent geometry space augmented with a discrete mesh-fidelity index. Candidate latent vectors are initialised with 30 Latin hypercube

samples to gain information on the design space. Once the objective function f has been evaluated, objectives are modelled with a `SingleTaskGP`, fitted by maximising the exact marginal log-likelihood. At each Bayesian optimisation iteration, an Expected Improvement (EI) acquisition function is constructed and, for every admissible fidelity, optimised over the latent dimensions. The (noise-free) Expected Improvement (EI) is

$$\text{EI}(z, s) = (\mu(z, s) - f^+) \Phi\left(\frac{\mu(z, s) - f^+}{\sigma(z, s)}\right) + \sigma(z, s) \phi\left(\frac{\mu(z, s) - f^+}{\sigma(z, s)}\right) \quad (7)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative density function and probability density function of the standard normal distribution, respectively.

$$\hat{c}(s) = \frac{1}{n_s} \sum_{i:s_i=s} c_i, \quad (8)$$

where c_i is the measured wall-clock time of evaluation i and n_s . The cost-adjusted acquisition function used to select the next best candidate to evaluate is then defined by:

$$\alpha_{\text{cost}}(z, s) = \frac{\text{EI}(z, s)}{\hat{c}(s)}, \quad \hat{c}(s) = \frac{1}{n_s} \sum_{i:s_i=s} c_i \quad (9)$$

where c_i is the measured wall-clock time of evaluation i and n_s . The next best point is then selected by maximising the cost-adjusted acquisition function. A maximum budget and iterations are defined such that if one condition is met, the optimisation stops. Forty Bayesian optimisation evaluations and a time budget of 18 hours were set using 32 CPUs, 12 of which were dedicated to the MPI solver for the particle balance equations.

5. CFD Setup

This study investigates the 2-D behaviour of paracetamol in a Newtonian fluid (water) to predict the PSD using a coupled CFD-PBE framework. A constant inlet velocity is used to ensure a laminar flow regime, stable supersaturation profiles and reproducible crystal growth behaviour. This is done to minimise temporal fluctuations in mixing, which directly contribute to secondary nucleation and create non-uniform PSDs [20]. This also reduces computational cost and ensures numerical stability in the CFD-PBE solver, allowing for repeated evaluations within the optimisation loop. Reactor geometries were generated using the VAE-based parameterisation with fixed surface areas. Straight pipe sections were appended at the inlet and outlet to promote fully developed flow as seen in Figure 4. They were then tested under these flow conditions to identify shapes that minimise the PSD and create a uniform-sized product for manufacturing.

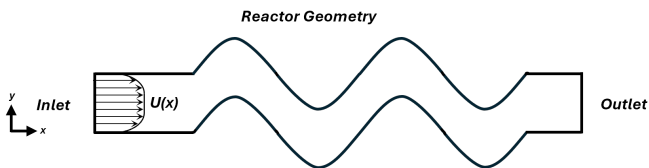


Figure 4: Nominal reactor geometry used for baseline performance studies

5.1. Governing Equations

The CFD simulation follows a two-step solving regime within OpenFOAM. First, `buoyantPimpleFoam` [21] is used for the hydrodynamics regime by solving the continuity (10), Navier Stokes (11) and scalar transport equations (12):

$$\nabla \cdot \mathbf{u} = 0 \quad (10)$$

$$\rho \frac{D\mathbf{u}}{Dt} = \nabla \cdot \mu (\nabla \mathbf{u} + \nabla \mathbf{u}^T) + \rho \mathbf{g} - \nabla p \quad (11)$$

$$\frac{DC}{Dt} = D\nabla^2 C \quad (12)$$

where $\nabla \cdot \mathbf{u}$ is the velocity evaluated over the stream-wise and span-wise coordinates, $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla$ is the total material derivative, where $\mathbf{u}$ is the fluid vector velocity in two dimensions.

The simulation is run for a sufficient time to allow the velocity field to converge ($\sim 3s$). The flow regime then switches to a PBE solver called `pbeTransportFoam` from the OpenQBMM suite [22]. This solver freezes the hydrodynamics data at the final timestep and uses it as a basis for evaluating the PBE. This approach is adopted because crystal evolution occurs on significantly longer timescales than flow convergence, allowing the hydrodynamics to be treated as a quasi-steady state whilst maintaining accurate PSD predictions.

5.2. Crystal Kinetics

Crystallisation behaviour in the reactor is governed by the solubility of paracetamol and the resulting degree of supersaturation. A solubility curve defines three operational regions, stable, metastable and labile (supersaturated), shown in Figure 5. These regions are determined by the degree of saturation, which is defined as the maximum amount of solute a solvent can dissolve at a given temperature, $c^*(T)$. It is possible to achieve supersaturation, $S = c/c^* > 1$, by cooling the solution or altering its concentration.

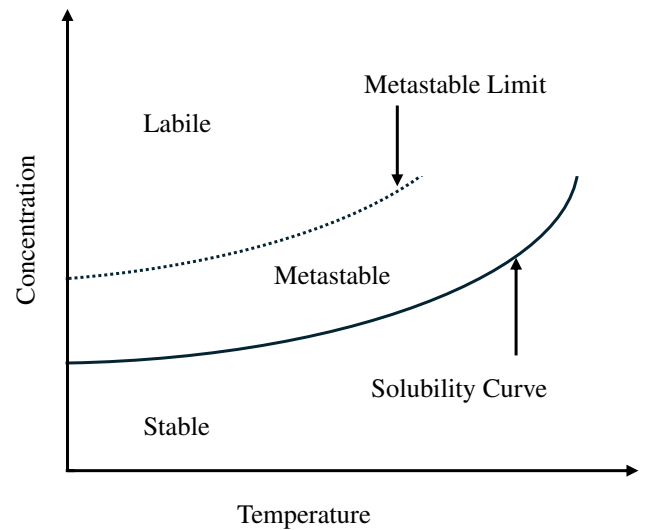


Figure 5: Generic solubility curve

At this point, a solution can hold more solute than theoretically possible, which is extremely thermodynamically unfavourable and leads to rapid, uncontrolled primary nucleation. Operating within the metastable zone (MSZ) promotes controlled crystal growth with minimal spontaneous nucleation and is therefore essential for producing a narrow PSD [23].

In the CFD-PBE framework used in this study, the solubility curve provides the basis for calculating growth rates and determining when nucleation occurs. Since supersaturation is strongly influenced by local hydrodynamics, coupling the solubility behaviour with the CFD field is crucial for accurately predicting the PSD.

5.3. Population Balance Modelling

The governing PBE equation describes the spatial and temporal evolution of a discrete population of droplets, bubbles or crystals. It is an evolution of the number density function (NDF), $n(\xi, \mathbf{x}, t)$, for which the internal coordinate ξ is selected to be particle length (L). The PBE incorporates discontinuous phenomena such as nucleation, aggregation, breakage and continuous phenomena like diffusion and convection and is defined as [4]:

$$\begin{aligned} \frac{\partial n(\xi, \mathbf{x}, t)}{\partial t} + \nabla_{\mathbf{x}} \cdot [n(\xi, \mathbf{x}, t) \mathbf{U}] - \nabla_{\mathbf{x}} \cdot [\Gamma \nabla_{\mathbf{x}} n(\xi, \mathbf{x}, t)] + \nabla_{\xi} \cdot [G(\xi) n(\xi, \mathbf{x}, t)] \\ = \bar{B}^a(\xi, \mathbf{x}, t) - \bar{D}^a(\xi, \mathbf{x}, t) + \bar{B}^b(\xi, \mathbf{x}, t) - \bar{D}^b(\xi, \mathbf{x}, t) + \mathcal{N}(\xi, \mathbf{x}, t) \end{aligned} \quad (13)$$

where $\mathbf{U}$ is the velocity of the carrier fluid, Γ is the diffusivity, $G(\xi)$ is the growth rate, $\bar{B}^a(\xi, \mathbf{x}, t)$, $\bar{D}^a(\xi, \mathbf{x}, t)$, $\bar{B}^b(\xi, \mathbf{x}, t)$ and $\bar{D}^b(\xi, \mathbf{x}, t)$ are the rates of change of n due to birth and death in the aggregation and breakage processes respectively and $\mathcal{N}(\xi, \mathbf{x}, t)$ is the rate of change due to nucleation.

5.4. Extended Quadrature Method of Moments

Solving the NDF is extremely computationally expensive, and subsequently, transport equations of a finite set of moments are solved to find an approximate solution to the PBE. In a univariate system, these moments are defined as:

$$M_k(\mathbf{x}, t) = \int_0^{+\infty} n(\xi, \mathbf{x}, t) \xi^k d\xi. \quad (14)$$

The initial moment set was constructed to represent a monodisperse, preseeded solution with a characteristic crystal size of $2 \mu\text{m}$, consistent with early stage crystallisation studies [24] [25]. The significance of each moment and the value they take in this study are outlined in Table 1.

Table 1: Moment Parameters

Parameter	Meaning	Value	Units
M_0	Number density	3.125×10^{12}	m^{-3}
M_1	Length density	6.25×10^6	m^{-2}
M_2	Surface-area density	12.5	m^{-1}
M_3	Volume density	2.5×10^{-5}	–
M_4	Kurtosis	5.0×10^{-11}	m

The transport equation for a k th order moment is obtained by multiplying Equation 13 by ξ^k and integrating over the entire domain $[0, +\infty]$. The transport equation is defined as:

$$\begin{aligned} \frac{\partial M_k(\mathbf{x}, t)}{\partial t} + \nabla_{\mathbf{x}} \cdot [M_k(\mathbf{x}, t) \mathbf{U}] - \nabla_{\mathbf{x}} \cdot [\Gamma \nabla_{\mathbf{x}} M_k(\mathbf{x}, t)] + \bar{G}_k(\mathbf{x}, t) \\ = \bar{B}_k^a(\mathbf{x}, t) - \bar{D}_k^a(\mathbf{x}, t) + \bar{B}_k^b(\mathbf{x}, t) - \bar{D}_k^b(\mathbf{x}, t) + \bar{\mathcal{N}}_k(\mathbf{x}, t). \end{aligned} \quad (15)$$

The moment transport equations contain source terms like growth, aggregation, breakage and nucleation. These terms require the complete NDF to be known, but the method of moments only transports a finite set of moments, causing a closure problem. Consequently, the NDF is approximated from the transported moments using a weighted sum of non-negative kernel density functions (KDFs) $\delta_{\sigma}(\xi, \xi_{\alpha})$:

$$n(\xi) \approx p_N(\xi) = \sum_{\alpha=1}^N w_{\alpha} \delta_{\sigma}(\xi, \xi_{\alpha}), \quad (16)$$

where N is the number of KDFs, coefficients w_{α} and ξ_{α} are the non-negative weights of each KDF and the corresponding primary quadrature abscissae respectively.

The KDF is chosen such that when $\sigma \rightarrow 0$, the KDF $\delta_{\sigma}(\xi, \xi_{\alpha})$ tends to the Dirac delta function $\delta(\xi - \xi_{\alpha})$, matching the behaviour of the quadrature method of moments. The parameters, $(w_{\alpha}, \xi_{\alpha})_{\alpha=1, \dots, N}$ and σ , are then computed from the transported moment set $(M_k)_{k=0}^{2N}$ using a moment inversion algorithm so that the reconstructed NDF has exactly the same moments as the original distribution in Equation 16.

Log-normal KDFs are employed in OpenFOAM because their support on $\mathbb{R}^+$ ensures physically realisable particle sizes within the reconstructed NDF:

$$\delta_{\sigma}(\xi, \mu) = \frac{1}{\xi \sigma \sqrt{2\pi}} e^{-\frac{(\ln \xi - \mu)^2}{2\sigma^2}}, \quad \xi, \mu, \sigma \in \mathbb{R}^+. \quad (17)$$

Once the NDF is reconstructed, it is used to evaluate the integrals in the source terms of the moment transport equations. The NDF is still not able to be computed analytically and therefore a secondary quadrature with weights $w_{\alpha\gamma}$, and abscissae, $\xi_{\alpha\gamma}$, is required. These parameters are determined by forming the Jacobi matrix associated with the log-normal KDF and solving the corresponding eigenvalue problem. Only the Stieltjes-Wigert quadrature guarantees the correct preservation of the moments of the NDF and is therefore used within OpenQBMM [4].

This work retains only the growth source term in the PBE. Operating within the MSZ ensures that primary nucleation is negligible, while the laminar flow regime minimises shear induced aggregation and breakage. As a result, crystal growth is the predominant mechanism governing the evolution of the PSD in the system. The growth source term describes the rate of change of moments due to molecular growth and is defined as:

$$\bar{G}_k(\mathbf{x}, t) = -k \int_0^{\infty} \xi^{k-1} G(\xi) n(\xi, \mathbf{x}, t) d\xi. \quad (18)$$

Using the extended quadrature approximation of the NDF, the growth term becomes:

$$\bar{G}_k(\mathbf{x}, t) \approx -k \sum_{\alpha=1}^N w_{\alpha} \sum_{\gamma=1}^{N_{\alpha}} w_{\alpha\gamma} \xi_{\alpha\gamma}^{k-1} G(\xi_{\alpha\gamma}, \mathbf{x}, t) \quad (19)$$

Expressing these higher order terms as a weighted sum converts intractable integrals into computable, algebraic expressions, ensuring practical closure and numerical accuracy whilst significantly reducing the computational cost. Simulations were run for 10 s using the pbeTransportFoam solver.

5.5. Assumptions

The main assumptions are stated below to reduce computational complexity for the benefit of the optimisation algorithm:

- The reactor walls are rigid and impose no-slip and no-penetration boundary conditions at the walls, $u|_{y=0} = u|_{y=H} = 0$.
- The reactor is maintained at constant temperature, $T = 293.15 \text{ K}$.
- The fluid is assumed incompressible and therefore its density remains constant throughout the domain, $\rho = 998 \text{ kg m}^{-3}$ [26].
- The solution is pre-seeded. Nucleation, aggregation, breakage and solute consumption due to crystal growth are all neglected. Growth rate is constant, $G = 5.55 \cdot 10^{-6} \text{ m s}^{-1}$ [27].

5.6. Mesh Sensitivity Analysis

A mesh sensitivity analysis was conducted to evaluate how mesh fidelity influences the maximum velocity at the outlet of the nominal reactor geometry shown in Figure 4. The flow was simulated for three seconds using buoyantPimpleFoam on five mesh qualities of increasing resolution, allowing the velocity to converge. This analysis demonstrates how sensitive the results are to mesh fidelity and the motivation for using variable mesh sizes in the MFBO framework. A 10% increase in velocity was observed when increasing the mesh fidelity from 4,000 cells to 12,250 after the velocity field had converged at 3 s, as shown in Figure 6.

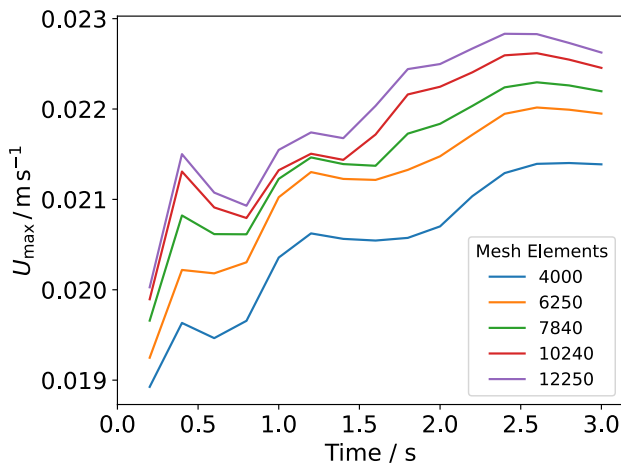


Figure 6: Maximum Outlet velocities for varying mesh fidelities

The corresponding computational time only increased by threefold from 45 s to 130 s as seen in Table 2. These mesh qualities increased the accuracy by a significant margin while keeping the computational within a reasonable range and therefore it was deemed a feasible domain to operate within for the optimisation algorithm.

Table 2: Mesh Selection Parameters

Mesh Size (x,y)	Computational Time (s)	Max Courant Number [-]
200 × 20	45	0.273768
250 × 25	55	0.349996
280 × 28	72	0.402034
320 × 32	99	0.467905
350 × 35	130	0.516482

As this is a reactor operating in the laminar regime, the inlet velocity was set to a fixed value of 0.01 m s^{-1} . This was verified by calculating the Reynolds Number at the inlet:

$$Re = \frac{\rho u d}{\mu} = \frac{998 \cdot 0.01 \cdot 0.01}{10^{-3}} = 99.8 \quad (20)$$

Water properties were obtained at the reactor operating temperature, where ρ is the density, u is the inlet velocity, d is the inlet diameter and μ is the dynamic viscosity. The Reynolds number is sufficiently below the threshold for turbulent flow < 2000 , confirming the original assumption of laminar flow.

The Courant Number was also monitored to ensure numerical stability in every simulation. It is defined as:

$$C = \frac{u \Delta t}{\Delta x} \quad (21)$$

where u is the velocity in each mesh cell, Δt is the time step and Δx is the mesh length. A conservative timestep value was selected based on simulations with the nominal geometry to ensure the Courant number remained well below 1. This provided a sufficient numerical stability margin for more complex geometries, and a nominal value of $\Delta t = 0.001 \text{ s}$ was chosen. After velocity convergence is reached, the flow field is frozen, fixing the Courant number for the entire PBE stage.

6. Objective Function

To guide the MFBO algorithm, the PSD must be translated into a quantitative performance metric. The objective function is derived from this distribution and provides a measure of how effectively a reactor geometry promotes uniform crystal growth.

6.1. Crystal Size Analysis

The PSD is constructed from the volume-weighted mean diameter, also known as the De Brouckere mean diameter. This is computed in each mesh cell using the the 3rd and 4th moments, M_3 and M_4 , as follows:

$$d_{43}(m) = \frac{M_4}{M_3} \quad (22)$$

These cell values are used to generate the PSD as shown in Figure 7. To obtain a smooth representation of the underlying distribution, a kernel density estimate is also computed. The KDE helps identify the first local minimum between the primary and secondary modes of the distribution. This is done to improve the fitting of the Tanks in Series Model in Section 6.3.

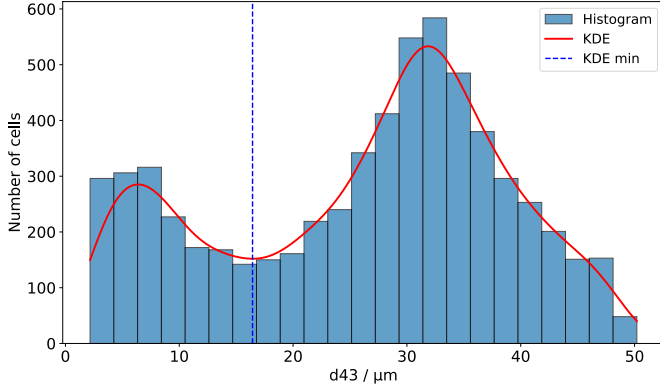


Figure 7: Example PSD curve fitting to determine the first local minima

6.2. Data Fitting

In the MFBO framework, the mesh fidelity varies between simulations and therefore so does the number of data points in the PSD. To ensure a consistent and data-driven analysis of the distribution across all geometries, the bin width is determined using the Freedman-Diaconis rule. This method selects the bin width, h , using the interquartile range (IQR) and number of data points, n , making it particularly robust for the irregular non-Gaussian distributions encountered in crystallisation. The bin width, h , and corresponding number of bins, k , are calculated as follows:

$$h = 2 \frac{\text{IQR}}{n^{1/3}}, \quad k = \left\lceil \frac{\max(x) - \min(x)}{h} \right\rceil \quad (23)$$

6.3. Tanks in Series Approximation

The tanks in series (TIS) model is used to quantify the uniformity of the PSD. It was originally developed to describe residence time distributions, but it is used in this study for its effectiveness in capturing unimodal distributions. It assumes N perfectly mixed continuous stirred reactors (CSTRs) connected in series. In the limit $N \rightarrow \infty$, the resulting distribution becomes progressively narrower and taller. This behaviour is desirable within the crystallisation process as it corresponds to a more uniform paracetamol crystal size. It is defined as:

$$\hat{E}_i(N) = \frac{N(N\theta_i)^{N-1}}{(N-1)!} e^{-N\theta_i}, \quad i = 1, \dots, d, \quad (24)$$

$$\text{MSE}(N) = \frac{1}{d} \sum_{i=1}^d (E_i - \hat{E}_i(N))^2 \quad (25)$$

where θ_i is the dimensionless particle size, calculated by dividing each measurement by the mean, and E_i is the normalised

bin density. The optimal value of N is obtained by minimising the mean squared error (MSE) between the dimensionless length and the TiS model. Once N is computed, it is used within the objective function for the optimisation algorithm, with larger values indicating a narrower and more uniform PSD.

7. Results

The MFBO algorithm was configured to maximise the value of N , resulting in a narrower and uniform PSD, using mesh fidelity as a cost variable. An initialisation step of 30 iterations using Latin hypercube sampling at the lowest fidelity was used to explore the design space. This facilitated rapid and computationally cheap evaluations, as shown for the first 30 evaluations in Figure 9.i. As the optimisation progressed, promising regions of the latent space were identified, and these were allocated a larger fraction of the remaining cost. This can be seen as the average evaluation cost increases from iterations 0 to 40, with the most expensive iterations occurring at the end of the algorithm, in a confined region of geometrically similar reactors. This allows the algorithm to search broadly first, using low-cost evaluations, and then evaluate the most promising geometries as accurately as possible.

The corresponding evolution of the objective value over the iterations is shown in Figure 8. The nominal geometry was evaluated early with a low N value of 11.82. As the design space was explored more thoroughly, the values of N generally increased, with small fluctuations occurring when new regions of the latent space were investigated. At iteration 1, an N value of 35.34 was identified, demonstrating a clear improvement compared to the nominal case. At the end of the optimisation, a reactor geometry was identified with an N value of 134.03, more than an order of magnitude larger than the nominal geometry. It should be noted that global optimality cannot be guaranteed for such a high-dimensional problem without gradient-based methods. However, these results show that the MFBO framework can efficiently explore the latent space using cheap, coarse meshes before honing in on high-performing candidates that generate geometries producing narrow, nearly unimodal PSDs. This represents a substantial enhancement over the nominal design.

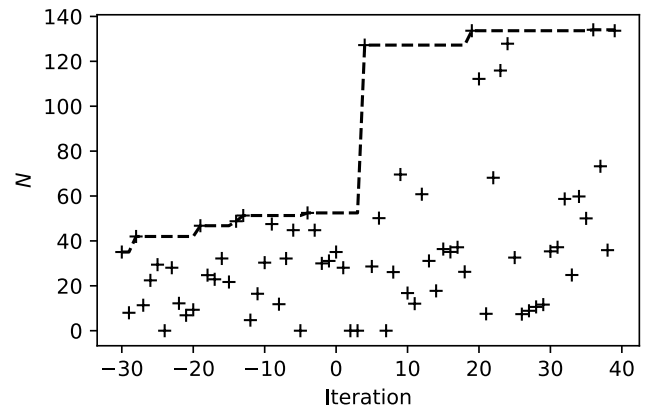


Figure 8: Iteration vs N across the optimisation algorithm

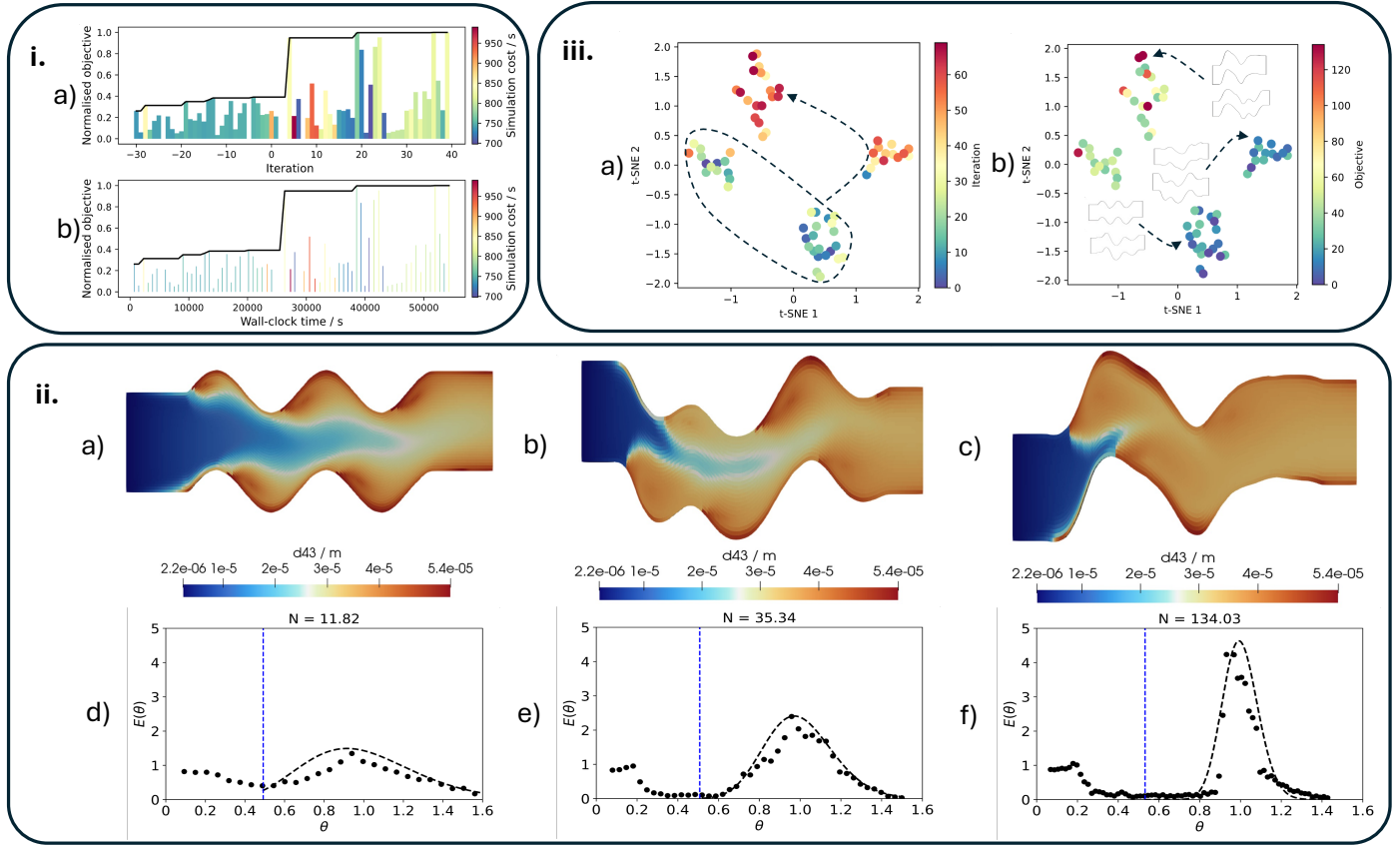


Figure 9: Analysis of PSD and optimisation performance: i) shows MFBO normalised optimisation objective with computational costs at each a) iteration and b) wall-clock time. ii) shows the performance comparison of volume-averaged mean particle size d_{43} in the a) nominal geometry (250x25), b) iteration 31 (280x28), and c) optimal solution (320x32) and the corresponding Tanks in series curve fitting for d) $N = 11.82$, e) $N = 35.34$, and f) $N = 134.03$. iii) shows a t-SNE analysis of the data generated throughout the optimisation in the design parameter space Z for a) iterations and b) objective values. iii.b) shows the distinct feature of the reactor geometries found at each cluster.

Figure 9iii.a) and b) visualise how the MFBO algorithm explores the learnt latent space with the t-SNE embedding of all evaluated geometries. Figure 9iii.a) illustrates how the algorithm progressively moves away from the initialisation dataset towards new regions of the latent space. Figure 9iii.b) highlights how these clusters share geometrical similarities, with the highest objective values occurring at higher iteration values. One geometry was selected from each cluster to examine their physical properties, as shown in Figure 9 ii.

In general, the largest crystals form near the walls where local velocities are lowest because of the no-slip boundary condition, and consequently, the residence times are the highest. In contrast, the smallest crystals are seen near the inlet and high velocity regions, where fluid elements pass through the reactor much more quickly. It should be noted that in the current model, when a layer of crystals forms at the walls, the no-slip condition is not re-applied at the crystal-fluid interface. The hydrodynamics therefore follow the underlying wall geometry. This simplification may overestimate residence time distributions near the wall, but it does not affect the qualitative trends discussed. In the nominal reactor, high-growth regions are found in the pockets formed by the sinusoidal waves, whilst the low-growth regions extend over a large part of the downstream channel. The

intermediate and optimal geometries are both irregular shapes with pronounced corrugations, but they differ significantly in their inlet shaping. In the optimal configuration, the channel bends upwards immediately and introduces a pronounced pinch point, creating a decelerated region with high residence times and enhanced radial mixing. Consequently, the velocity field and corresponding crystal growth become much more homogeneous. This leads to the narrowest and sharpest peak in the PSD, giving the largest value of N , representing the best crystallisation reactor design within the explored space.

8. Discussion

This study has highlighted the effectiveness of the 2D crystallisation reactor optimisation framework in identifying reactor geometries that promote highly uniform PSDs. Through the use of a VAE, this research enabled the learning of reactor parameterisation into a continuous space, improving on previous research. Having a continuous, well-informed latent space improved optimisation performance, enabling the identification of geometric properties to enhance PSDs. However, several natural extensions would increase both the physical realism of the model and the complexity of the framework. Modelling geometries in 3D would allow the capture of secondary flows, recirculation zones and inlet and outlet manifolds. All of these strongly

influence the mixing and residence time distributions. This transition would substantially increase the computational cost of each CFD-PBE evaluation, requiring a cost Gaussian Process to better understand changes in computational cost at individual fidelities. Taking a constrained MFBO approach on pressure drop would allow other reactor performance constraints to be met. Additionally, the crystallisation kinetics have been greatly simplified due to the time constraints of the project and to focus efforts on setting up the entire framework first. To increase the accuracy of the PSD, nucleation, aggregation and breakage should be re-enabled. Introducing a supersaturation calculator that couples local concentration and temperature would enable a dynamic evaluation of the MSZ rather than assuming a fixed operation in this region. Integrating these into the current model would be challenging, but it would form the next steps toward a physically accurate, 3D optimisation framework capable of guiding the design of optimal crystallisation reactors.

To keep such extended models tractable, future work could focus on integrating more advanced machine learning surrogate models with existing large image datasets into the optimisation pipeline. Diffusion models could be used to learn richer cross-sectional geometries, providing an intelligent generative knowledge of reactor shapes that can be efficiently sampled during optimisation. Transfer learning would allow the utilisation of already trained diffusion models that can encode and decode images effectively at a reduced computational expense.

9. Data Availability

All code used for the framework associated with this article can be found in the following repository: <https://github.com/Computational-Design-Intelligence/Crystallisation-Reactor>

References

- [1] Paracetamol market – industry report. Accessed: 5/12/2025.
- [2] Meifen Jiang and Xiong-Wei Ni. Effects of water and temperature on reaction mechanism and crystal properties in a reactive crystallization of paracetamol. *Chemical Engineering and Processing - Process Intensification*, 131:20–26, 2018.
- [3] Leila Keshavarz, Mahboubeh Pishnamazi, U.B. Rao Khandavilli, Saeed Shirazian, Maurice N. Collins, Gavin M. Walker, and Patrick J. Frawley. Tailoring crystal size distributions for product performance, compaction of paracetamol. *Arabian Journal of Chemistry*, 14(4):103089, 2021.
- [4] A. Passalacqua, F. Laurent, E. Madadi-Kandjani, J.C. Heylmun, and R.O. Fox. An open-source quadrature-based population balance solver for openfoam. *Chemical Engineering Science*, 176:306–318, 2018.
- [5] Onur Can Boy, Idelfonso BR Nogueira, Ulderico Di Caprio, and Mumin Enis Leblebici. Ai-aided process intensification of structures. *Chemical Engineering and Processing-Process Intensification*, page 110406, 2025.
- [6] Seyyed Amirreza Abdollahi, Ali Basem, As'ad Alizadeh, Dheyaa J. Jasim, Mohsen Ahmed, Abbas J. Sultan, Seyyed Faramarz Ranjbar, and Hamid Maleki. Combining artificial intelligence and computational fluid dynamics for optimal design of laterally perforated finned heat sinks. *Results in Engineering*, 21:102002, 2024.
- [7] Seongeon Park, Jonggeol Na, Minjun Kim, and Jong Min Lee. Multi-objective bayesian optimization of chemical reactor design using computational fluid dynamics. *Computers Chemical Engineering*, 119:25–37, 2018.
- [8] Tom Savage, Nausheen Basha, Jonathan McDonough, James Krassowski, Omar Matar, and Ehecatl Antonio del Rio Chanona. Machine learning-assisted discovery of flow reactor designs. *Nature Chemical Engineering*, 1(8):522–531, 2024.
- [9] Mahtab Izadi, Masoud Rahimi, and Reza Beigzadeh. Evaluation of micromixing in helically coiled microreactors using artificial intelligence approaches. *Chemical Engineering Journal*, 356:570–579, 2019.
- [10] Michael Mansour, Katharina Zähringer, Krishna D.P. Nigam, Dominique Thévenin, and Gábor Janiga. Multi-objective optimization of liquid-liquid mixing in helical pipes using genetic algorithms coupled with computational fluid dynamics. *Chemical Engineering Journal*, 391:123570, 2020.
- [11] Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.
- [12] Samuel Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space. In *Proceedings of the 20th SIGNLL conference on computational natural language learning*, pages 10–21, 2016.
- [13] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 world wide web conference*, pages 689–698, 2018.
- [14] Ishaan Gulrajani, Kundan Kumar, Faruk Ahmed, Adrien Ali Taiga, Francesco Visin, David Vazquez, and Aaron Courville. Pixelvae: A latent variable model for natural images. *arXiv preprint arXiv:1611.05013*, 2016.
- [15] Diederik P Kingma, Max Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.
- [16] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. *Advances in neural information processing systems*, 29, 2016.
- [17] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pages 1278–1286. PMLR, 2014.
- [18] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*, 2017.
- [19] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M. Jorge Cardoso. *Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations*, page 240–248. Springer International Publishing, 2017.
- [20] Shailesh Agrawal and Anthony Paterson. Secondary nucleation: Mechanisms and models. *Chemical Engineering Communications*, 202, 01 2015.
- [21] OpenFOAM Foundation. *buoyantPimpleFoam — OpenFOAM Solver for Buoyant Flow*, 2025. Accessed: 11/12/2025.
- [22] OpenQBMM Community. *pbeTransportFoam – Verification and Validation of the One-Way Coupled PBE Solver*, 2024. Accessed: 10/12/2025.
- [23] Yan Zhang. 6 - modeling and simulation of batch and continuous crystallization processes. In Syed Ahmad Imtiaz, editor, *Modelling of Chemical Process Systems*, pages 181–202. Elsevier, 2023.
- [24] Javad Karimi Sabet, Cyrus Ghotbi, and Farid Dorkoosh. Application of response surface methodology for optimization of paracetamol particles formation by res method. *Journal of Nanomaterials*, 2012(1):340379, 2012.
- [25] Zoltan K. Nagy and Richard D. Braatz. Robust nonlinear model predictive control of batch processes. *AIChE Journal*, 49(7):1776–1786, 2003.
- [26] NIST Chemistry WebBook. Nist chemistry webbook: Water (thermophysical properties at 293.15 k), 2025. Accessed: 11/12/2025.
- [27] Niall A. Mitchell, Clifford T. Ó'Ciardhá, and Patrick J. Frawley. Estimation of the growth kinetics for the cooling crystallisation of paracetamol and ethanol solutions. *Journal of Crystal Growth*, 328(1):39–49, 2011.

Carbon Capture on Waste-to-Energy Plant

Fanqi Gong and Adam Zeng

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

This study develops and evaluates an integrated waste-to-energy (WTE) and post-combustion carbon capture system using Aspen Plus. The process begins with RDF drying and incineration to generate high-temperature flue gas and steam for a Rankine cycle, producing approximately 18.2MW of electricity. The flue gas is treated using amine-based CO₂ capture, with MEA selected as the base case solvent. Rate-based modelling was employed for absorber and stripper columns to accurately represent mass-transfer behaviour. Sensitivity analysis identified lean loading, solvent concentration, pressure and temperature as the dominant parameters influencing specific reboiler duty (SRB) and capture efficiency, leading to an optimised base-case value of 3.4 GJ/tCO₂. Advanced process configuration, including Cold Solvent Split (CSS) and Lean Vapour Recompression (LVR), were investigated to enhance thermal performance. CSS reduced cooling duty significantly, while LVR achieved a reboiler duty of 2.31 GJ/tCO₂ with higher compressor power and capital cost as its trade-off. A blended MEA/PZ solvent system was also assessed, demonstrating improved reaction kinetics, lower circulation rates, and favourable energy performance. Despite the higher unit cost of PZ, the blend achieved both CAPEX and OPEX reductions due to smaller equipment sizing and reduced regeneration demand.

1. Introduction

Growing concerns over climate change and the need to reduce greenhouse gas emissions are driving many sectors to reassess their environmental impact. At the same time, rapid urbanisation and population growth continue to generate large quantities of municipal and commercial waste, placing sustained pressure on existing waste-management systems. Traditional landfilling is increasingly constrained due to environmental regulations, rising landfill taxes, and long-term sustainability concerns[1]. These limitations have accelerated the adoption of Waste-to-Energy (WtE) incineration, which has become a key treatment route for residual waste and an important contributor to the UK's energy supply.

However, despite its advantages, WtE incineration produces a high level of CO₂ emissions due to the carbon-rich composition of residual municipal waste, and its fossil CO₂ intensity can exceed the modern gas-fired power plants[2]. As the UK moves toward net-zero emissions and prepares to introduce carbon pricing for WtE facilities, developing effective CO₂-mitigation strategies for this sector is becoming increasingly important.

This project focused on evaluating the feasibility of post-combustion CO₂ capture for a representative WtE plant. The study examined solvent performance under high-CO₂ conditions, beginning with MEA as the base case and extending to blended solvents to improve capture efficiency and energy performance. In addition to conventional solvent-based configurations, advanced process modifications were explored to identify opportunities for reducing reboiler duty and enhancing overall system efficiency. By combining baseline process modelling, detailed solvent analysis, and the evaluation of advanced process, this work aimed to build a clear and detailed technical understanding of post-combustion CO₂ System. An economic assessment was included to complement the technical study and highlight the wider implications of applying carbon capture technology to the WtE sector. Together, these analyses provided insight into how WtE facilities can be adapted to operate with lower CO₂ emissions while maintaining their essential role in waste management.

2. Background

2.1 MSW for power generation

Municipal solid waste (MSW) is a mixed fuel, made up of both biogenic material and fossil-derived products. In most cases, the residual waste stream is dominated by biodegradable items that mainly food scraps and garden waste. While paper, cardboard, and plastic packaging still make up a large part of what ends up in the bin. Smaller amounts of other combustibles such as textiles, rubber, and wood also contribute to the overall heating value of the waste. On the other hand, materials like metals, glass, and inert residues do not burn and simply dilute the fuel, reducing the energy that can be recovered. Refuse-derived fuel (RDF) is produced from municipal solid waste by removing non-combustible materials and processing the remaining combustible fraction into a more uniform, energy-rich feed suitable for Waste-to-Energy plants. Because MSW depends heavily on local consumption habits and waste-management practices, its composition can vary widely from one country to another. Table 3.1.1 summarises the types of materials typically found in UK waste used for power generation, with equivalent data from the Netherlands included for comparison. These differences highlight why MSW composition matters, it affects how much energy can be produced and how much CO₂ is released during combustion[3].

2.2 Solvent selection

MEA was selected as the base-case solvent because it is the most established and widely studied amine for post-combustion CO₂ capture. As a primary amine, MEA provides fast reaction kinetics, good absorption efficiency, and predictable regeneration behaviour, making it a reliable benchmark for evaluating capture performance. Its thermodynamic and kinetic properties are well documented in literature and fully supported by simulation tools, allowing accurate modelling of absorber–stripper behaviour and energy requirements. These qualities make MEA the standard reference solvent for CO₂-capture studies and an appropriate starting point for this investigation.

<u>RDF</u>	<u>PROXANAL</u>				<u>ULTANAL</u>					<u>LHV</u>
	<u>Moisture</u> <u>content</u>	<u>Volatile</u> <u>matter</u>	<u>Fixed</u> <u>carbon</u>	<u>Ash</u>	<u>C</u>	<u>H</u>	<u>O</u>	<u>N</u>	<u>S</u>	<u>MJ/kg</u>
<u>UK[4]</u>	<u>5.9</u>	<u>70.0</u>	<u>11.2</u>	<u>12.9</u>	<u>58.7</u>	<u>8.4</u>	<u>16.0</u>	<u>1</u>	<u>0.4</u>	<u>24.9</u>
<u>Netherlands[5]</u>	<u>10.8</u>	<u>66.5</u>	<u>8.6</u>	<u>14.1</u>	<u>54.6</u>	<u>8.37</u>	<u>21.61</u>	<u>0.91</u>	<u>0.41</u>	<u>22.6</u>

Table 3.1.1 RDF composition

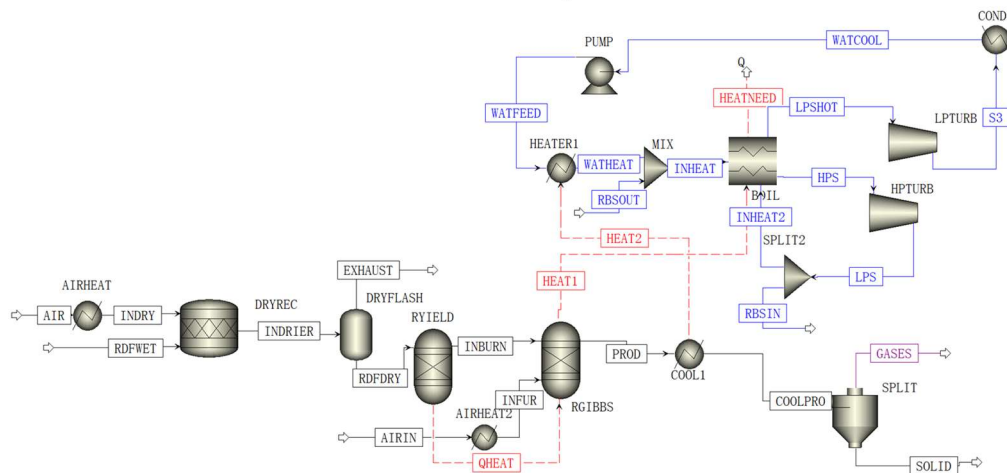


Figure 3.1.1 Waste-to-Energy Plant flowsheet

3. Power generation plant

3.1 Waste-to-energy plant

A WtE plant incinerating RDF to generate electricity was modelled in Aspen Plus. Figure 3.1.1 is the flowsheet of the simulation.

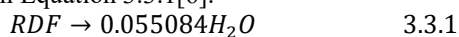
The simulation included 2 important processes, including combustion of RDF and Rankine cycle electricity generation.

3.2 RDF Feed

In the property section of Aspen, proximate analysis and ultimate analysis data are necessary for nonconventional fuels like RDF. The data is listed in Table 3.1.1. RDF of the UK was chosen for the simulation due to a larger LHV value, which means that it can release more energy while incinerating. A feed rate of 15,000kg/h was chosen for this study, corresponding to a mid-scale single incineration line.

3.3 Drying

As RDF usually has a relatively high moisture content, drying is essential to ensure efficient combustion. A RSTOIC reactor block was used to model the drying process by mixing the original RDF with hot air. The stoichiometry was set by assuming coal's behaviour in drying process, which is showed in Equation 3.3.1[6].



The RDF stream is then processed through a flash column to separate vapour from dry RDF.

3.4 Incineration

Dry RDF then went through 2 reactor blocks to simulate the incineration process. The first reactor block RYIELD does not consider chemical reactions. Instead, it decomposes nonconventional components into a set of conventional components by user-defined data. In this case, RDF needed to be converted into conventional components (C,H,O,N,S) in RYIELD block based on the ultimate analysis data. The stream outlet then underwent actual incineration reaction in RGIBBS. RGIBBS finds the equilibrium composition that minimise Gibbs free energy. Since Aspen determines the

equilibrium state purely on thermodynamics, there is no need to specify individual reactions. For satisfying overall energy balance, a heat stream was connected between RYIELD and RGIBBS. The outlet product stream of RGIBBS had a temperature of 1300°C, which was then cooled down to 200°C, providing 11.17MW heat to boiler inside the steam power cycle.

3.5 Flue gas and particulate separation

Post combustion gas-solid separation was carried out using a cyclone separator. The cyclone can provide a robust stage of particulate removal, which can significantly reduce the dust load before the flue gas enters downstream carbon capture process. This step can enhance overall system reliability by limiting erosion, fouling, and particulate carryover.

3.6 Steam cycle and power generation

Electricity generation in the waste-to-energy plant was carried out using Rankine cycle. High-pressure turbine with isentropic efficiency of 0.66 and mechanical efficiency of 0.98 and low-pressure turbine with isentropic efficiency of 0.8 and mechanical efficiency of 0.8 were used in the simulation[7].

The thermal energy released from RDF combustion was transferred to water in the boiler, which was modelled as a heat exchanger in Aspen to produce high-pressure superheated steam. This steam was expanded first through a high-pressure (HP) turbine, where a portion of the thermal energy was converted into mechanical work of 9.80MW. After partial expansion, the steam was returned to the boiler to restore its temperature and improve overall cycle efficiency. The steam was then being directed into a low-pressure (LP) turbine for further expansion. The LP turbine extracted the remaining useful energy, generating electrical output of 8.42MW This reduced the steam to a low-pressure state suitable for condensation. The exhaust steam then entered the condenser, where it was cooled to saturated liquid and pumped back to the boiler, completing the cycle. In this case, the waste-to-energy plant generated

approximately 18.2MW electricity, before integrating with the carbon capture process.

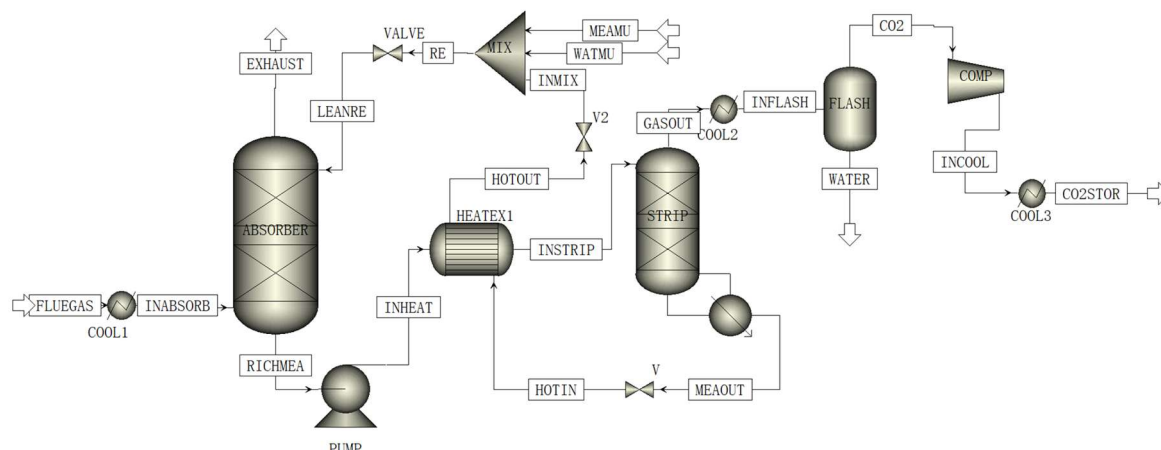
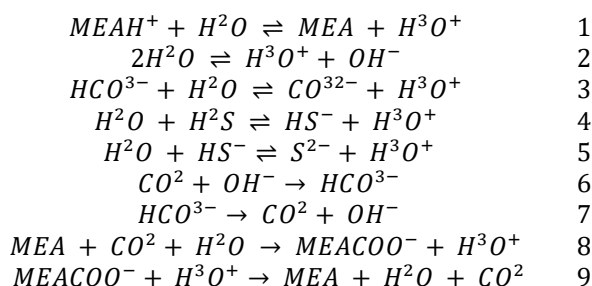


Figure 5.1.1 Carbon Capture Base Case Flowsheet

4. Carbon Capture Preliminary Study

4.1 MEA Solvent Characteristics

MEA is the most widely used and investigated solvent for post-combustion CO₂ capture and was selected to be the base case solvent. As a primary amine, MEA reacts rapidly with CO₂ to form carbamate, enabling high absorption rates even at low CO₂ partial pressures suitable for WTE flue gas. The reaction mechanism for $MEA - CO_2 - H_2O$ system is listed below.[8]



The formation and decomposition of bicarbonate is described by kinetic reactions 6 and 7. While reaction 8 and 9 represents the formation and hydrolysis of carbonate in absorber and stripper respectively. H_2S reaction is included as the flue gas contains the sulphur species.

4.2 Method Selection for Modelling

The ENRTL (Electrolyte NRTL) method was selected for the following modelling. ENRTL is designed for aqueous electrolyte systems and accurately represents the non-ideal liquid behaviour, ionic species, and reaction equilibria in $MEA - CO_2 - H_2O$ solutions. The comparison between experimental results and method prediction results for components can be found in the reference provided[8]. A brief comparison was made with ENRTL-RK, which applies the Redlich–Kwong EOS to the vapour phase for high-pressure CO_2 systems. Since the column operate at low pressure, non-ideal vapour is limited and the additional complexity is unnecessary. A detailed comparison between two methods is listed in the appendix.

The flowsheet of the carbon capture process is shown in Figure 5.1.1. Flue gas has chemical reaction with lean MEA in the absorber. The CO_2 rich MEA stream from the bottom goes through a heat exchanger to increase temperature. The rich MEA stream then undergoes the chemical reaction in reverse direction in the stripper to release CO_2 from the top, while regenerating the solvent in the bottom stream. Water is separated from the CO_2 stream in a flash column. The treated CO_2 is then compressed cooled for storage. The regenerated solvent is used to heat the rich MEA stream in the heat exchanger before being recycled back into the absorber with the make-up streams.

5.2 Lean loading analysis

An essential parameter in the modelling process is the lean loading, as it directly influences regeneration energy and CO₂ loading capacity. Sensitivity analysis in Figure 5.2.1 shows a consistent trend which is higher MEA mass fractions lead to lower reboiler duty. All tested compositions also display a U-shaped profile with respect to lean loading, indicating that each solvent has an optimal operating point. Across the range, a lean loading of 0.23 mol CO₂/mol MEA delivered the lowest energy demand and was selected for the simulations. MEA concentrations above 30% were not considered due to the corrosion concerns.[9]

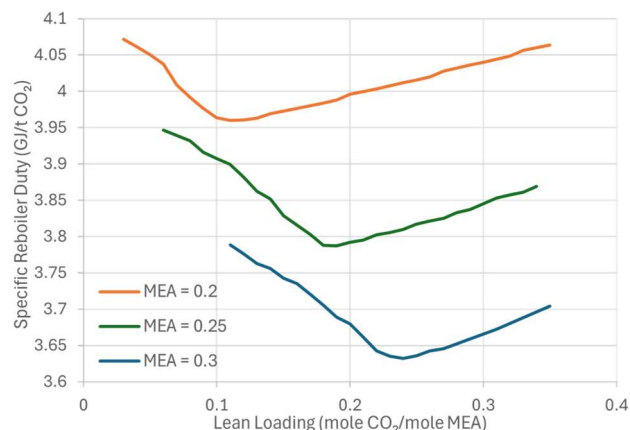


Figure 5.2.1 Specific Reboiler duty against Lean Loading

5.3 Flue gas

The flue gas generated by the power plant was processed in a carbon capture process. The composition of the flue gas is listed in Table 5.3.1

	Mass fraction
CO ₂	0.184
H ₂ O	0.078
N ₂	0.692
H ₂ , H ₂ S, HCl, SO ₂ , CO	0.04

Table 5.3.1 Flue Gas Composition

5.4 Cooler

The flue gas was cooled from 200°C to 35°C before entering the bottom stage of the absorber.

5.5 Absorber design

The absorber was designed as a packed column using Sulzer MELLAPAK™ 250Y due to large surface area for liquid–gas contact, low pressure drop, and highly predictable hydraulic behavior.

Lean MEA entered the top stage of the absorber at 48°C. These parameters were determined by sensitivity analysis which helped optimise the system. The dimensions of the column were set to ensure that it is hydraulic feasible. The column design of the absorber is shown in Table 5.5.1

Parameter	Specification
Internal Type	packed
Packing Type	MELLAPAK™
Dimension	250Y
Height	20 m
Diameter	5 m
Pressure	1 bar

Table 5.5.1 Absorber Design

5.6 Heat exchanger design

A counter-current heat exchanger was utilised to heat the rich MEA solvent to 105°C.

5.7 Stripper design

Same as the absorber, Sulzer MELLAPAK™ 250Y was also selected for the packed column of the stripper. Parameters including inlet temperature and pressure, column pressure, boil up ratio were determined by implementing sensitivity analysis, aiming to minimise reboiler duty and increase capture efficiency. The final specific reboiler duty was 3.4GJ/tCO₂. The column design of the stripper is shown in Table 5.7.1

Parameter	Specification
Internal Type	packed
Packing Type	MELLAPAK™
Dimension	250Y
Height	16 m
Diameter	2.4 m
Pressure	2 bar
Boil up ratio	0.1093

Table 5.7.1 Stripper Design

5.8 Flash column

Since CO₂ has strict storage requirements for purity.[10] The flash column setting to 25°C and 3bar was used to remove water from the CO₂.

5.9 CO₂ compression and storage

CO₂ needs to be at a very high pressure for transportation in pipelines. A compressor was implemented to pressurise the CO₂ stream to 150 bar[11]. Since the compressor caused huge temperature increase up to 491°C, a cooler was used to cool it down to 50°C.

5.10 Rate-based modelling

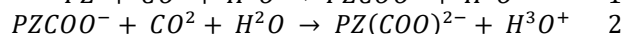
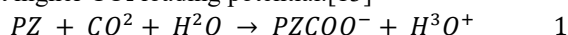
Rate-based modelling was used for column simulations including both absorber and stripper. It is preferred over equilibrium modelling when accuracy in predicting mass and

heat transfer is important, especially in amine-based carbon capture process.[12]

6. Blended Solvent Investigation (MEA + PZ)

6.1 Introduction to Blended Solvents

Blending amines is a widely used strategy to overcome the limitations of single-solvent systems such as MEA. The main motivation for blending is to combine the strengths of different amines to achieve faster CO₂ absorption, higher working capacity, and lower regeneration energy. Piperazine (PZ) is one of the most effective promoters for MEA because it exhibits very fast reaction kinetics and can form both mono- and di-carbamate species as shown in the reaction below, giving it higher CO₂ loading potential.[13]



When added in small quantities, PZ increases the overall reaction rate and reduces the required solvent circulation rate for a given capture target. As a result, blended MEA/PZ solvents can achieve higher capture efficiency with improved absorber performance compared to pure MEA.

6.2 Simulation Setup for MEA/PZ Blends

To clearly understand the behaviour of the blended solvent, the first and most important step was to investigate how different solvent compositions affect absorption performance and energy demand, because the composition directly affects CO₂ loading capacity, and regeneration behaviour, therefore the systematic testing of MEA–PZ ratios was essential.

Based on previous single-solvent optimisation work, MEA = 0.3 mass fraction was identified as the best-performing MEA concentration while respecting its corrosive nature. Consequently, the first series of tests kept MEA fixed at 0.3 while varying the PZ content to evaluate how increasing PZ affects capture efficiency and reboiler duty. Three investigated compositions are listed below, 30% MEA with 1%PZ, 5%PZ, 10%PZ

After establishing the trend within the fixed-MEA group, a second series of simulations was performed where the total amine concentration was kept constant value of 0.3 weight fraction. In this step, MEA and PZ were redistributed within the same total amine loading. 25%MEA/5% PZ ; 20%MEA/10% PZ ; 22.5 %MEA/7.5%PZ

6.3 Lean loading analysis (MEA+PZ)

To investigate the performance of each combination, lean loading analysis was carried out by varied lean loading for all MEA/PZ compositions while recording the change in specific reboiler duty. The first being analyse was the fixed MEA mass fraction groups.

Figure 6.3.1 illustrates the results of the sensitivity analysis. Across all solvent compositions, the specific reboiler duty decreases as lean loading increases from low values and reaches a minimum around 0.21 mol CO₂/mol amine, after that SRD starts to rise again. This produces the characteristic U-shaped trend observed in all three curves. Although MEA/PZ blends provide faster absorption kinetics, all tested mixtures show higher SRD than the single-solvent MEA baseline (3.5 GJ/tCO₂).

Among the tested solvent blends, the 30% MEA / 5% PZ formulation achieved the best performance, giving the lowest specific reboiler duty of approximately 4.27 GJ/tCO₂. In contrast, the 30% MEA / 10% PZ blend consistently

shows the highest SRD, suggesting that excessive PZ can actually lead to a poorer energy performance.

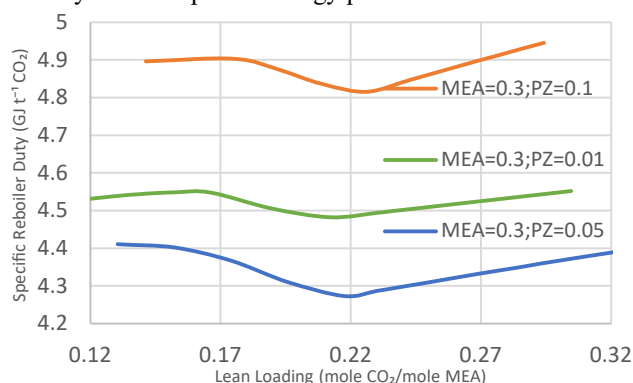


Figure 6.3.1 Specific Reboiler duty against Lean Loading

A similar lean loading analysis was carried out for the blends with fixed total amine content as shown in Figure 6.3.2, and all compositions exhibit the same characteristic U-shaped SRD trend, with an optimal lean loading close to 0.23 mol CO₂/mol amine. Compared to the base case, this group achieves better overall energy performance, with the lowest specific reboiler duty of approximately 3.25 GJ/tCO₂ at a lean loading of 0.227. Among the tested blends, 22.5% MEA / 7.5% PZ delivers the best performance, achieving the lowest SRD. In contrast, the 25% MEA / 5% PZ mixture shows the weakest performance within the group, indicating that the distribution of MEA and PZ amine content plays an important role in regeneration energy and does not follow a simple linear trend.

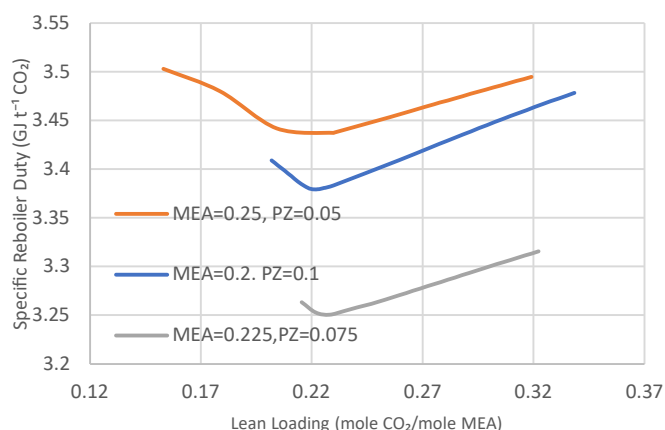


Figure 6.3.2 Specific Reboiler duty against Lean Loading

7. Sensitivity Analysis

Base case

7.1 Absorber analysis

7.1.1 Lean solvent flowrate

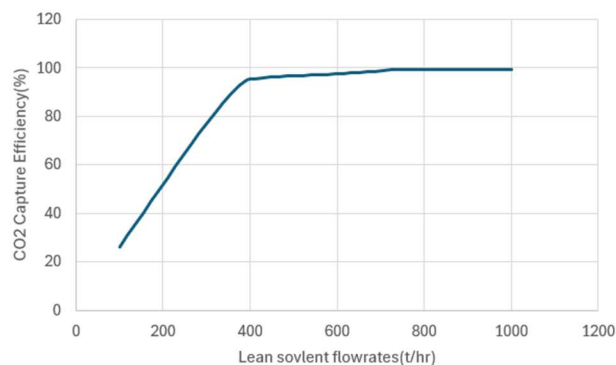


Figure 7.1.1 Capture Efficiency against Lean solvent flowrate

The absorber analysis aimed to identify operating conditions that can achieve the required capture efficiency while minimising solvent circulation, which is linked to the cost. Figure 6.3.2 indicates that a lean-solvent flow rate of approximately 400 tonnes/hr is needed to exceed 90% capture efficiency. Beyond this point, further increases in solvent flow provide only marginal improvement.

7.1.2 Lean solvent temperature

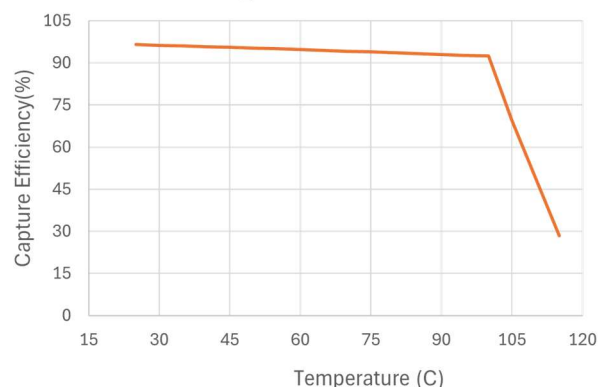


Figure 7.1.2 Capture Efficiency against lean solvent temperature

Figure 7.1.2 shows that lean-solvent temperature has a direct influence on capture efficiency, and the recycled stream should remain below 90 °C for reaching the capture efficiency required.

7.1.3 Flue gas temperature

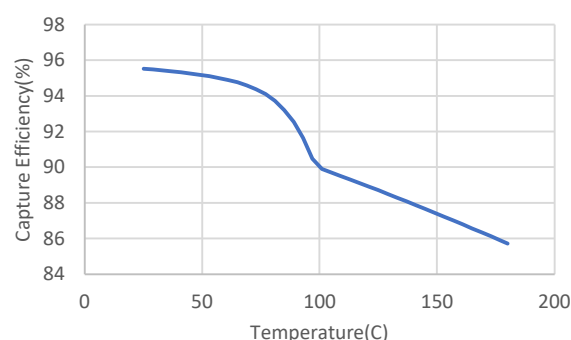


Figure 7.1.3 Capture Efficiency against flue gas temperature

Similarly, Figure 7.1.3 indicates that flue gas feed in temperature also affects the adsorption performance, a cooler was therefore installed to cool the flue gas to 35°C.

7.2 Stripper analysis

Stripper analysis was necessary for the design for minimising reboiler duty and making sure CO₂ was well separated from MEA to reach a high capture efficiency.

7.2.1 Inlet Temperature

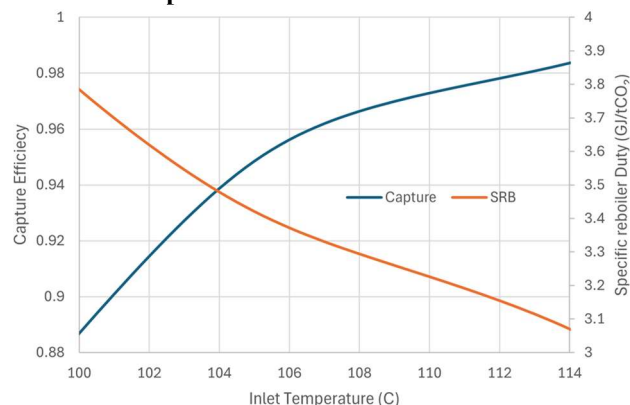


Figure 7.2.1 SRB & Capture Efficiency against Inlet Temperature

According to Figure 7.2.1, higher Inlet Temperature leads to better capture efficiency and lower reboiler duty. An inlet temperature of 105°C was determined considering a relatively low reboiler duty and the heating feasibility of the heat exchanger before the stripper.

7.2.2 Inlet Pressure

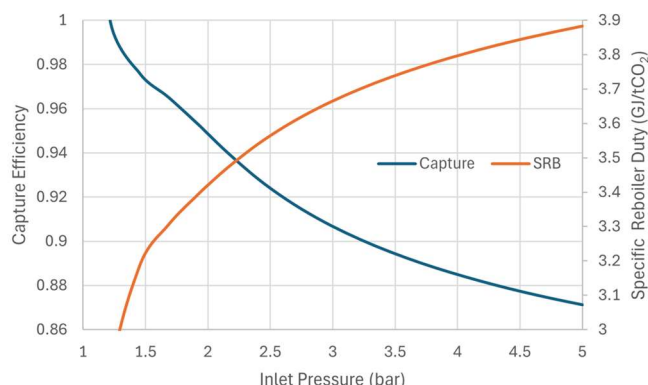


Figure 7.2.2 SRB & Capture Efficiency against Inlet Pressure

Based on Figure 7.2.2, it can be concluded that high inlet pressure results in smaller capture efficiency and higher reboiler duty. 2 bar was chosen to reach over 94% capture efficiency and moderate reboiler duty.

7.2.3 Boil up ratio

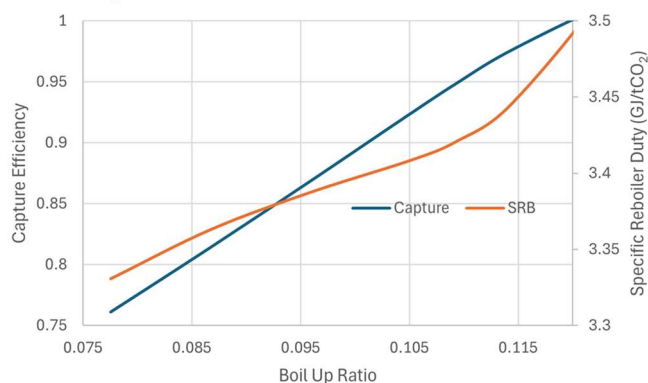


Figure 7.2.3 SRB & Capture Efficiency against Boil Up Ratio

Figure 7.2.3 illustrates that higher boil up ratio causes higher capture efficiency and higher reboiler duty. There is a trade-off between a good capture efficiency and a low reboiler duty. Thus, a boil up ratio of 0.1093 was chosen to balance both conditions.

7.2.4 Column pressure

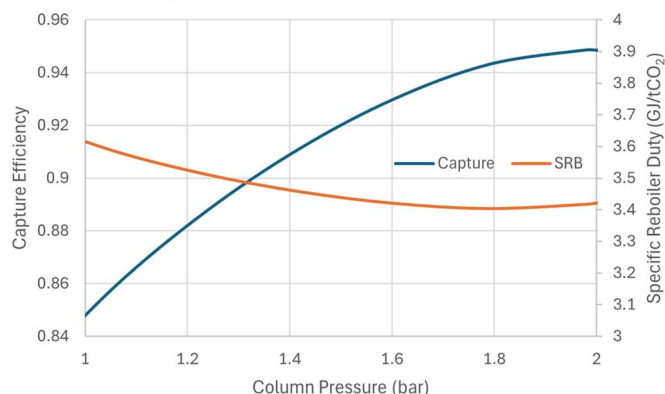


Figure 7.2.4 SRB & Capture Efficiency against Column Pressure

Internal Column pressure is also a very important parameter that can affect simulation results. Figure 7.2.3 shows that higher column pressure results in higher capture efficiency and smaller reboiler duty. Both trends reach a plateau

between 1.8 bar and 2 bar. 2 bar was selected as the optimal value, while matching the inlet stream pressure.

Blended solvent case

7.3 Absorber analysis

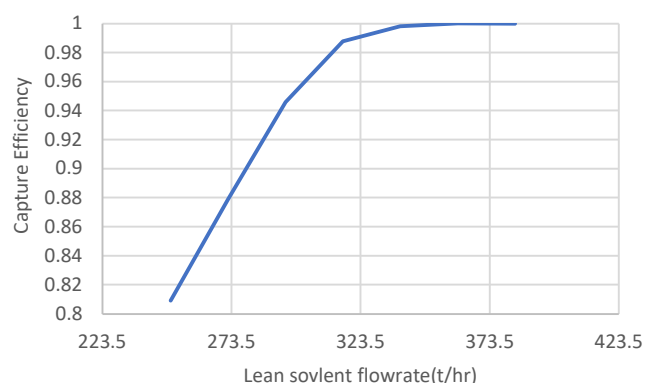


Figure 7.3.1 Lean solvent flowrate vs. capture Efficiency at 30% MEA/5%PZ

Figure 7.3.1 demonstrates how the addition of PZ to the solvent influences capture efficiency in the blended system. As expected, increasing the lean solvent flow rate leads to higher capture efficiency for all cases. However, the introduction of PZ significantly enhances absorption performance: achieving more than 90% capture now requires only approximately 280 t/h of lean solvent, compared with 400 t/h for the MEA-only base case. This improvement arises from the chemistry of PZ, which can form both mono- and di-carbamate species, thereby increasing CO₂ loading capacity (as discussed in Section 4). A more detailed comparison of solvent compositions and their associated performance trends is provided in the appendix.

7.4 Stripper analysis

7.4.1 Inlet Temperature

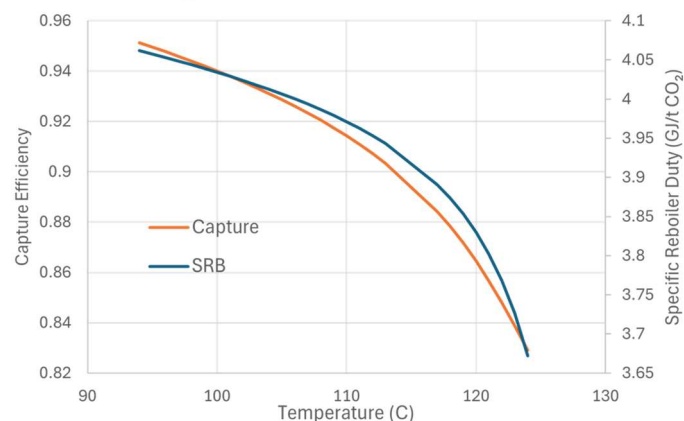


Figure 7.4.1 SRB & Capture Efficiency against Temperature

The effect of the stripper feed temperature was also examined to identify potential performance improvements. As with previous cases, increasing the inlet temperature decreases the reboiler duty. Unlike the single-solvent MEA system, the blended solvent exhibits an opposite response in terms of capture performance shown by Figure 7.4.1. Higher feed temperatures now result in lower capture efficiency. This introduces a clear trade-off, where improving energy performance through elevated feed temperature simultaneously decrease CO₂ removal. Therefore, identifying an operating temperature that satisfies both the energy and capture efficiency targets becomes more challenging for the blended solvent.

7.4.2 Inlet Pressure

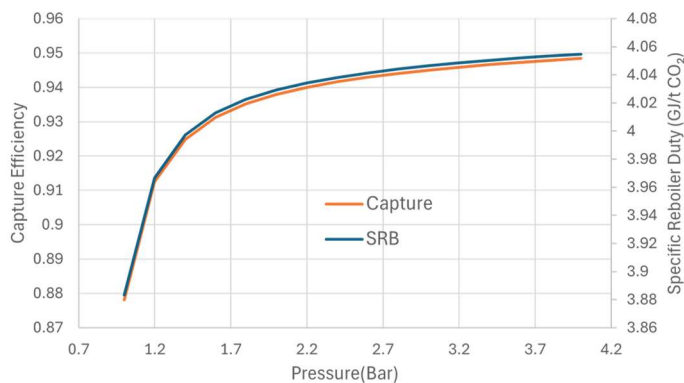


Figure 7.4.2 SRB & Capture Efficiency against Pressure

The relations between stripper inlet pressure and capture efficiency, specific reboiler duty is shown below by Figure 7.4.2. In contrast to the single-solvent base case, increasing the stripper pressure now results in higher capture efficiency, indicating that the blended MEA/PZ solvent benefits from the enhanced solvent–gas contact and improved reaction driving forces at elevated pressures. The effect on energy performance remains consistent with the base case, so pressure can improve capture performance for the blended solvent, but with a higher cost of regeneration energy.

8 Advanced Process Modification

8.1 Cold solvent Split (CSS)

Cold Solvent Split is an energy-saving modification by splitting the MEA stream into 2 fractions [14]. One portion is still pre-heated by the heat exchanger before entering the stripper. Another portion goes directly into the top stage of the column as a cold stream with a temperature of 35°C. The purpose of this design is to reduce the temperature of the top of the stripper and reduce the cooling duty of the cooler after the stripper. The cooling duty of this exact cooler is reduced from -28.02 GJ/hr to -13.12 GJ/hr, which was decreased by more than 50%.

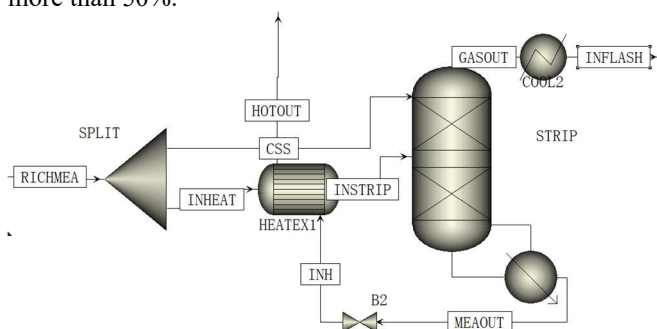


Figure 8.1.1 CSS configuration

8.2 Lean Vapour Recompression (LVR)

Lean vapour recompression is an advanced process modification applied to the bottom product of the stripper. The lean solvent is routed through a flash unit to generate a vapour stream, which is then compressed and returned to the lower section of the stripper. Compression raises the temperature of the recycled vapour, allowing it to supply additional thermal energy for solvent regeneration and thereby significantly reducing the reboiler duty. The final optimised specific reboiler duty is 2.31 GJ/tCO₂ which is much lower than base case that support the theory of LVR[15].

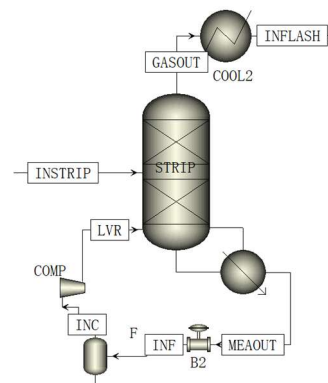


Figure 8.2.1 LVR configuration

To evaluate the performance of the LVR configuration, the behaviour of the two key units which is the adiabatic flash and the compressor was examined using sensitivity analysis.

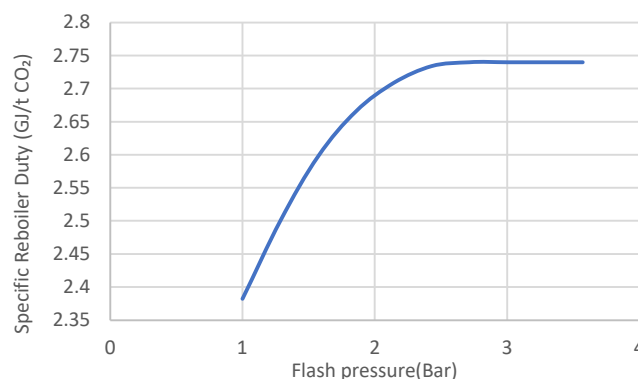


Figure 8.2.1 SRB against Flash operation pressure

The flash-pressure study in Figure 8.2.1 shows a clear positive correlation between flash operating pressure and stripper reboiler duty. At higher flash pressures, the amount of vapour generated in the flash decreases, which reduces the enthalpy available for recompression. As a result, the recycled vapour provides less heating to the stripper, causing the reboiler duty to increase. Beyond roughly 3 bar, this effect reaches the limit because the vapour fraction becomes sufficiently low that further pressure increases have little additional impact on column energy balance.

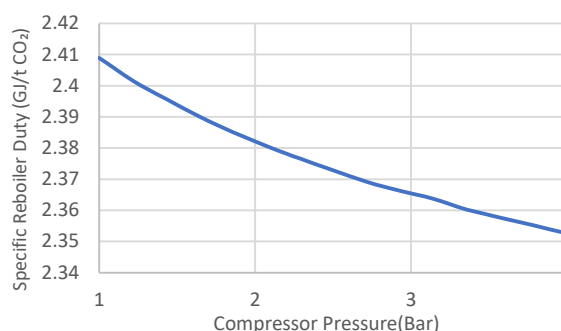


Figure 8.2.3 SRB against Compressor discharge pressure

In contrast, increasing the compressor discharge pressure produces a beneficial reduction in reboiler duty as shown in Figure 8.2.3. Higher discharge pressures lead to higher outlet temperatures due to compression heating, and the hotter recycled vapour supplies more thermal energy to the lower section of the stripper. This improves solvent regeneration efficiency. However, the improvement comes with a cost that is higher discharge pressures significantly increase the compressor power requirement, introducing a trade-off between thermal savings and electrical consumption.

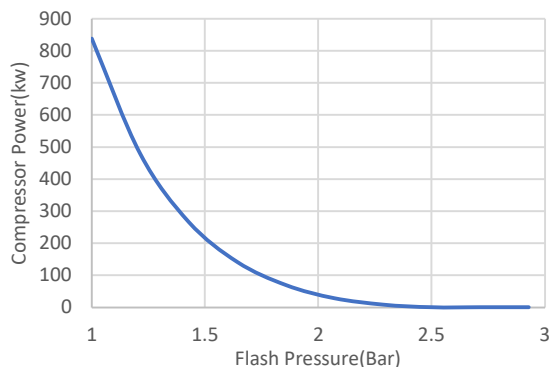


Figure 8.2.4 Compressor power required against Flash pressure

The flash pressure directly influences compressor power demand as shown in Figure 8.2.4 shown, higher flash pressures reduce the compression ratio and therefore lower electricity usage. At the same time, increasing flash pressure raises reboiler duty, meaning that an optimal balance must be identified to minimise total energy consumption.

9 Economic analysis

To better compare all the different cases and their economic viability. Economic analysis was done by calculating capital expenditure, operational expenditure and raw material cost

9.1 Capital Expenditure (CAPEX)

The economic evaluation in this study followed a structured cost-estimation workflow based on the methodology presented in Douglas' Conceptual Design of Chemical Processes.[16] For each major unit operation, the base equipment cost was first calculated using the sizing correlations and cost equations provided in the book. These base costs were then converted to total installed costs by applying an appropriate Lang factor, which accounts for installation, piping, auxiliary systems, and indirect construction expenses. A Lang Factor of 5 was chosen.[17] Finally, all costs were updated to current economic conditions using the Chemical Engineering Plant Cost Index (CEPCI), allowing the original Douglas correlations that was derived from historical price data to be accurately transformed to present-day values. $CEPCI_{2024}$ is 800, while $CEPCI_{1968}$ is 113.6, provided by Chemical Engineering Magazine.[18]

$$CAPEX = \frac{CEPCI_{2024}}{CEPCI_{1968}} \times LF \times EC \quad (1)$$

$$EC_{compressor} = 517.5 \times bhp^{0.82} \times (2.11 + Fc) \quad (2)$$

$$EC_{heat\ exchange} = 101.3 \times A^{0.65} \times (2.29 + Fc) \quad (3)$$

$$EC_{column} = 101.9 \times D^{1.066} \times H^{0.802} \times (2.18 + Fc) \quad (4)$$

$$A = \frac{Q}{U \Delta T_{LM}} \quad (5)$$

The equations above are the Guthrie Corrections used for calculating equipment cost. Equation (2) calculates the cost of the gas compressor with the horsepower value directly read off from simulation results. Equation (3) calculates the cost of heat exchangers, including the coolers, the reboiler in the stripper and the heat exchanger. Equation (5) calculates the area (A). Equation (4) calculates the cost of the columns including absorber and stripper. Fc represents the correction factor that is related to material, configuration type and pressure. CAPEX was then calculated by multiplying CEPCI and Lang Factor.

	CAPEX \$
Absorber	11.81M
Stripper	7.24M
Heat exchanger	12.33M

cooler	3.44M
pump	0.48M
flash	0.18M
compressor	47.67M

Table 9.1.1 Base Case CAPEX

According to Table 9.1.1, The overall Capital expenditure is approximately \$83M. The cost of compressor is about \$47.7M, occupying over half of the CAPEX.

9.2 CAPEX comparison with other cases

Economic analysis was also conducted for all alternative cases, including LVR, intercooling, CSS and the blended solvent case (MEA/PZ).

	LVR \$	CSS \$	MEA/PZ \$
Absorber	11.81M	11.81M	11.97M
Stripper	7.05M	7.29M	8.21M
Heat exchanger	4.51M	6.87M	6.82M
cooler	3.15M	2.92M	3.20M
pump	0.48M	0.48M	0.48M
flash	0.58M	0.20M	0.19M
compressor	60.33M	47.75M	48.68M
sum	87.90M	77.12M	79.55M

Table 9.2.1 Alternative Case CAPEX

Table 9.2.1 illustrates that CSS can result in lower capital expenditure while LVR results in higher capital cost. Even if it can reduce reboiler duty, thus reduce stripper cost, LVR consists of extra compressor and flash column. Compressor is the most expensive equipment in the whole CCS system, causing huge increase in overall cost.

The blended solvent case saves approximately \$3M comparing to base case. The heat exchanger cost is reduced by 44.7%. This is caused by a smaller overall stream flowrate, resulting in smaller heat exchanger area required.

9.3 Operational Expenditure (OPEX)

The operating expenditure of this study was evaluated based on utility consumption. Heat duties and shaft work of the units were directly obtained from simulation results. The unit price of each utility is showed in Table 9.3.1

	Unit price
Cooling water	0.325 \$/m ³
steam	0.040 \$/kwh
electricity	0.085 \$/kwh

Table 9.3.1 Utility unit price[19]

Cooling water is required for coolers, resulting in a cost of \$2.39M/year. Steam is supplied to the reboiler, with a cost of \$6.50M/year. This cost can be eliminated by utilising excess steam leaving the WTE system. Electricity is used to power pumps and compressors, with a cost of \$1.88M/year. This can also be covered by directly utilising electricity generated from the power plant. The annual operating cost is \$10.77M/year, if the cancellation is not considered.

9.4 OPEX comparison with other cases

	LVR \$	CSS \$	MEA/PZ \$
Cooling water	2.10M	1.93M	2.19M
steam	4.39M	6.61M	6.22M
electricity	2.25M	1.88M	1.89M
sum	8.74M	10.42M	10.30M

Table 9.4.1 Alternative Case OPEX

Table 9.4.1 shows the operating cost of LVR, CSS and blended solvent case. LVR reduces the steam cost by 32.5%

but increases the electricity cost by 19.7%. This is caused by the extra compressor implemented in the system. CSS reduces the cooling water cost by 19.2%. While the blended solvent case saves OPEX by 4.1% because of the reduction in reboiler duty.

9.5 Raw Material Cost

Since there are make-up streams in the system. Raw material cost needs to be evaluated. The comparison is mainly done between the base case and the blended solvent case. The unit price of each component is shown in Table 9.5.1

	Price(\$/kg)
MEA	1.49
PZ	8.00
water	0.003

Table 9.5.1 Raw Material price [20]

Overall, the base case requires \$2.93M/yr for make-up streams. The blended solvent case requires \$2.74M/yr, despite using PZ which is a much more expensive solvent. This means that the amount of money saved by reducing the amount of inlet stream is larger than the extra cost by the new solvent.

9.6 Revenue Generation

The revenue is directly related to the electricity generated from the WTE plant. Cancelling out the amount of electricity used by compressor and pump. As a result, the revenue before integrating with CCS is \$12.37M, which is then reduced to \$10.50M considering the whole WtE-CCS system.

10 Discussion

The results across all configurations provide a comprehensive view of how solvent chemistry, column operation, and process modifications influence the technical and economic performance of post-combustion CO₂ capture for WtE flue gas. The MEA base case offers a reliable benchmark, achieving a 94% capture efficiency and a specific reboiler duty (SRD) of 3.4 GJ/tCO₂, values consistent with industrial experience for 30wt% MEA systems operating with rate-based modelling. The sensitivity analyses highlight the dominant role of lean loading, temperature, and pressure in determining system performance. Most importantly, the U-shaped SRD–lean-loading curve indicates competing thermodynamic effects. At low lean loading, the solvent enters the stripper with very little CO₂, causing the VLE to shift unfavourably and increasing water evaporation. This raises the reboiler duty. At high lean loading, the solvent carries more CO₂, requiring a larger circulation rate to maintain the capture target. The increased flow raises the sensible heating requirement, which will increase the reboiler duty.[9] The optimum operating point (0.23 mol mol⁻¹) reflects the balance between cyclic capacity and regeneration energy.

The MEA/PZ blended solvent delivers markedly different behaviour due to its enhanced reaction kinetics and the ability of PZ to form both mono- and di-carbamate species.[13] The absorber performance improves substantially, requiring only 280t/h of lean solvent to achieve 94% capture, compared with 400t/h for MEA. This reflects both higher mass-transfer coefficients and a larger effective cyclic capacity. However, the stripper performance shows more complex trends. Although blended solvents generally achieve lower circulation rates, the SRD does not decrease proportionally unless the MEA–PZ ratio is carefully optimised. The best-performing blend (22.5% MEA / 7.5% PZ) achieved an SRD

of 3.25 GJ/tCO₂, confirming that modest PZ concentrations enhance performance, whereas excessive PZ reduces mass transfer, thereby worsening energy performance.

The advanced configurations provide further insight into system-level optimisation. Cold Solvent Split (CSS) successfully reduces the cooling duty by more than 50% by lowering the overhead temperature in the stripper, consistent with literature on thermal integration strategies[15]. However, its limited impact on reboiler duty indicates that the main regeneration load is still dominated by solvent desorption energy rather than rectification requirements at the top of the column. Lean Vapour Recompression (LVR), by contrast, directly reduces the reboiler load by using overheated compressed vapour to provide part of the regeneration heat. The observed SRD of 2.31GJ/tCO₂ aligns with previously reported reductions of reboiler duty in LVR-enhanced systems[14][15]. Nevertheless, the significant compressor power requirement and large equipment size result in a substantial CAPEX increase. This highlights the classical trade-off between thermal and electrical energy: improvements in regeneration heat demand may come at the expense of additional mechanical work.

Integrating the technical and economic outcomes reveals that no single configuration is universally superior. LVR achieves the strongest thermal improvement but imposes the highest capital cost, making it potentially attractive only when steam is expensive or waste heat is unavailable. CSS provides moderate energy savings with relatively low capital penalty and may therefore be the most practical enhancement for industrial retrofits. The MEA/PZ blend offers a balanced improvement, reducing both CAPEX and OPEX due to smaller solvent flow and reduced equipment sizing, even though PZ itself is more expensive. This aligns with trends in emerging commercial solvent systems that prioritise high cyclic capacity and low circulation rates.

11 Conclusion

This study successfully modelled and evaluated a full waste-to-energy (WtE) plant generating 18.2MW electricity and post-combustion carbon capture system using Aspen Plus, integrating RDF combustion, Rankine cycle power generation, and amine-based CO₂ absorption. The base MEA process provided a solid benchmark, achieving 94% capture efficiency with a specific reboiler duty of 3.4GJ/tCO₂. Sensitivity analyses reaffirmed the strong dependence of absorber and stripper performance on operating temperature, pressure, lean loading, and boil-up ratio, all of which guided the optimal design parameters selected for the base case.

Advanced designs were also simulated to seek improvements on the base case. LVR reduces the specific reboiler duty to 2.31GJ/tCO₂. CSS reduces the cooling duty of cooler after the stripper to -13.12GJ/hr. The blended solvent case also provides useful results on reducing overall stream inlet and reboiler heat duty (3.25GJ/tCO₂) by introducing another solvent piperazine (PZ) to mix with MEA.

Overall, comparing all configurations shows that while LVR delivers the greatest energy reduction, its economic penalty is substantial. CSS offers moderate technical improvement at lower cost. The MEA/PZ blended solvent achieves meaningful reductions in energy demand and equipment size with favourable overall economics. Therefore, for a WtE-CCS system of the scale investigated, Future work may expand on pilot-scale validation, dynamic simulation, and

optimisation under variable operating conditions to further strengthen these conclusions.

Reference

- [1] Energy recovery for residual waste A carbon based modelling approach. (2014). Available at: <https://nsip-documents.planninginspectorate.gov.uk/published-documents/EN010093-000942-GLA%20Deadline%207%20Submission%20-%20Appendix%20B%20Defra%20carbon%20modelling%20report.pdf>
- [2] Gov.uk (2014). *Energy from waste A guide to the debate*. [online] Available at: <https://assets.publishing.service.gov.uk/media/5a7c77ade5274a559005a113/pb14130-energy-waste-201402.pdf>.
- [3] Department for Environment Food & Rural Affairs (2021). Waste Management Plan for England. [online] Available at: <https://assets.publishing.service.gov.uk/media/60103f71d3bf7f05bc42d294/waste-management-plan-for-england-2021.pdf>.
- [4] Materazzi, M., Lettieri, P., Mazzei, L., Taylor, R. and Chapman, C. (2015). Fate and behavior of inorganic constituents of RDF in a two stage fluid bed-plasma gasification plant. *Fuel*, 150, pp.473–485. doi:<https://doi.org/10.1016/j.fuel.2015.02.059>.
- [5] Bosmans, A., Vanderreydt, I., Geysen, D. and Helsen, L. (2013). The crucial role of Waste-to-Energy technologies in enhanced landfill mining: a technology review. *Journal of Cleaner Production*, 55, pp.10–23. doi:<https://doi.org/10.1016/j.jclepro.2012.05.032>.
- [6] Aspen Technology, Inc. (2003) *Aspen Plus 12.1: Getting Started – Modeling Processes with Solids*. Cambridge, MA: Aspen Technology, Inc.
- [7] Anutoiu, S., Dosa, I. and Petrilean, D.C. (2021). Steam turbine efficiency assessment, first step towards sustainable electricity production. *MATEC Web of Conferences*, [online] 342, p.04007. doi:<https://doi.org/10.1051/mateconf/202134204007>.
- [8] Aspen Technology, Inc. (2012). Rate-Based Model of the CO₂ Capture Process by MEA using Aspen Plus. AspenTech Technical Documentation.
- [9] Soltani, S.M., Fennell, P.S. and Mac Dowell, N. (2017). A parametric study of CO₂ capture from gas-fired power plants using monoethanolamine (MEA). *International Journal of Greenhouse Gas Control*, 63, pp.321–328. doi:<https://doi.org/10.1016/j.ijggc.2017.06.001>.
- [10] de Visser, E., Hendriks, C., Barrio, M., Mølnvik, M.J., de Koeijer, G., Liljemark, S. and Le Gallo, Y. (2008). Dynamis CO₂ quality recommendations. *International Journal of Greenhouse Gas Control*, 2(4), pp.478–484. doi:<https://doi.org/10.1016/j.ijggc.2008.04.006>.
- [11] Intergovernmental Panel on Climate Change (IPCC) (2005) *IPCC Special Report on Carbon Dioxide Capture and Storage*. Cambridge: Cambridge University Press. Available at: <https://www.ipcc.ch/report/carbon-dioxide-capture-and-storage/>
- [12] Zhang, M. and Guo, Y. (2013). Rate based modeling of absorption and regeneration for CO₂ capture by aqueous ammonia solution. *Applied Energy*, 111, pp.142–152. doi:<https://doi.org/10.1016/j.apenergy.2013.04.074>.
- [13] Zhang, T., Yu, Y. and Zhang, Z. (2018). An interactive chemical enhancement of CO₂ capture in the MEA/PZ/AMP/DEA binary solutions. *International Journal of Greenhouse Gas Control*, [online] 74, pp.119–129. doi:<https://doi.org/10.1016/j.ijggc.2018.04.023>.
- [14] Jung, J., Jeong, Y.S., Lee, U., Lim, Y. and Han, C. (2015). New Configuration of the CO₂ Capture Process Using Aqueous Monoethanolamine for Coal-Fired Power Plants. *Industrial & Engineering Chemistry Research*, 54(15), pp.3865–3878. doi:<https://doi.org/10.1021/ie504784p>.
- [15] Zhang, W., Chen, J., Luo, X. and Wang, M. (2017). Modelling and process analysis of post-combustion carbon capture with the blend of 2-amino-2-methyl-1-propanol and piperazine. *International Journal of Greenhouse Gas Control*, 63, pp.37–46. doi:<https://doi.org/10.1016/j.ijggc.2017.04.018>.
- [16] Douglas, J. M. (1988). *Conceptual design of chemical processes*. McGraw-Hill.
- [17] Husain Bahzad, Grazia Leonzio, Soltani, S.M., Abdullah Al Jasmi, Davies, W.G., Fennell, P.S. and Ali, N. (2025). Techno-economic analyses of a novel hydrogen production process via chemical looping water splitting, integrated with sorption enhanced water gas shift. *International Journal of Hydrogen Energy*, 188, pp.151996–151996. doi:<https://doi.org/10.1016/j.ijhydene.2025.151996>.
- [18] Chemical Engineering Magazine (2024) *CEPCI report: Economic indicators, September 2024*. Available at: https://www.chemengonline.com/wp-content/uploads/2024/08/19_CHE_0924_EconInd_p45.pdf
- [19] Department for Business, Energy & Industrial Strategy (2022) *AECOM next-generation carbon capture technology: techno-economic analysis*. Available at: <https://assets.publishing.service.gov.uk/media/6294923ce9e07039e31b777/aecom-next-gen-carbon-capture-technology-technoeconomic-analysis.pdf>
- [20] Otitoju, O., Oko, E. and Wang, M. (2021). Technical and economic performance assessment of post-combustion carbon capture using piperazine for large scale natural gas combined cycle power plants through process simulation. *Applied Energy*, 292, p.116893. doi:<https://doi.org/10.1016/j.apenergy.2021.116893>.

Assessing Flexibility in the Haber-Bosch Process Under Variable Throughput

Amirthan Jeyavathanan and Rami Salame

Department of Chemical Engineering, Imperial College London, U.K.

Abstract Ammonia synthesis plays a key role in global chemical production and is increasingly viewed as a vital facilitator of hydrogen storage and renewable energy transport. However, adapting the traditionally rigid Haber-Bosch process to operate under fluctuating electrolyser-derived hydrogen remains a substantial challenge. An integrated steady-state Aspen Plus model was developed to examine the feasibility of a conventional Haber-Bosch loop facilitating green ammonia production without radical redesign. The model integrates cryogenic air separation, a multistage compressor train with off-design efficiency correlations, an optimised reactor system that uses Temkin-Pyzhev kinetics, fixed-UA heat exchangers, and a recycle-purge loop. Using reactor inlet temperatures as the main control variable, throughput was varied from 25% to 135% of design flow. The analysis revealed that mechanical and thermal constraints significantly limit flexibility. Reactor inlet temperatures significantly drop at low throughput due to excessive cooling in the heat-integration network, and at high throughput, heating and cooling duties surpass exchanger UA limits. The lower feasibility bound is ultimately determined by compressor turndown, with the upper bound determined by heat exchanger performance capacity. The feasibility analysis indicates a steady-state operating range of 65-135%. Therefore, this study lays the foundation for designing future green ammonia plants by establishing that flexibility is technically feasible but dependent on improved heat-integration techniques and compressor configuration.

Introduction

The global transition to renewable energy has intensified research into methods of energy storage. Beyond its industrial role as a precursor for fertilisers and explosives, ammonia is gathering significant interest as a promising medium for the storage and transportation of renewably derived hydrogen. Furthermore, traditional ammonia synthesis via the Haber-Bosch process accounts for around 1.8% of global carbon dioxide emissions, primarily due to the production of hydrogen through steam methane reforming (The Royal Society, 2020). Thus, a twofold challenge exists: to decarbonise the existing ammonia industry and to find a sustainable energy vector. Green ammonia emerges as a critical solution to meeting both challenges by acting as a hydrogen vector and providing a sustainable production alternative for one of the world's 'big four' chemical products, hence helping society achieve global net-zero goals.

However, the integration of renewable hydrogen production with the Haber-Bosch process poses a major obstacle: electrolysers powered by intermittent renewable energy resources, such as wind and solar, generate fluctuating hydrogen outflows. Conventional Haber-Bosch plants operate at a continuous steady-state with minimal flexibility, introducing uncertainty about how upstream fluctuations will propagate through the system. Also, an extensive understanding exists for the conventional steady-state Haber-Bosch process, whereas flexible operation allows for further investigation into the feasible operating range of green ammonia production with upstream variance.

This paper intends to explore further into flexible operation through the development of a steady-state model and a subsequent feasibility analysis. A steady-state flowsheet is constructed on Aspen Plus, integrating nitrogen production via cryogenic distillation of air, hydrogen sourced from liquid alkaline electrolyser performance data, a multi-stage compressor train with interstage cooling, and a three-reactor synthesis loop utilising a wüstite catalyst with recycle and purge. The constructed model is used to explore the feasible

operating range under variable feed rates based on equipment turndown ratios and the theoretical flexibility of the system. The resulting feasibility analysis illustrates the range of viable hydrogen throughputs and the corresponding operability of the process, outlining the extent to which variables must remain flexible to accommodate fluctuating hydrogen throughputs, hence providing a foundation for future work on flexible green ammonia production.

Background

The traditional Haber-Bosch process is intended for large-scale, steady-state operation with continuous hydrogen production from steam methane reforming (Smith et al., 2020). Green hydrogen produced from water electrolysis presents a challenge due to its reliance on intermittent renewable energy sources such as wind and solar (Galan et al., 2024). Unlike the steady hydrogen flow obtained from fossil-based sources, an electrolyser's output fluctuates since the electrical current feeding the electrolyser controls the production rate of hydrogen (Queiroz et al., 2024). Currently, less than 1% of the world's hydrogen production is classified as green, which can be attributed to its higher cost, approximately three times more expensive than grey hydrogen (Queiroz et al., 2024). As green hydrogen becomes more cost-effective and the demand for green ammonia increases, understanding the flexibility limits of ammonia synthesis becomes critical to determine whether the process can operate reliably under fluctuating feed conditions.

The conventional method for green hydrogen synthesis is water electrolysis, where electricity from renewable sources splits water into hydrogen and oxygen. Commercial electrolysers today include several variants, namely liquid alkaline (LA), proton exchange membrane (PEM), solid oxide electrolyser cell (SOEC), and the novel anion exchange membrane (AEM). LA units, operating at approximately 60-90°C, are the oldest and cheapest of current electrolyser technology at scale, with their capital cost ranging from \$800 to \$1,000 per kW, while also consuming 50-55 kWh of electricity per

kg of hydrogen (Jaeger and deBiasi, 2023). PEM electrolyzers utilise a solid polymer electrolyte and platinum catalysts, yielding similar energy efficiency but at a much higher cost in the vicinity of \$1,400 to \$1,700 per kW (Jaeger and deBiasi, 2023). The PEM design allows for high-pressure hydrogen output and a rapid dynamic response, enabling integration with variable renewable power (Hubert et al., 2024). SOEC electrolyzers, a new and emerging technology, operate at around 800°C; nevertheless, the industrial heat emitted can be used to generate 40 kWh/kg, improving energy efficiency (Jaeger and deBiasi, 2023). AEM electrolysis, which embeds alkaline chemistry in a solid membrane, is another emerging approach that runs on pure water, yielding lower capital expenditure. However, it currently lags in durability and performance in comparison to other viable methods of green hydrogen synthesis. (Hubert et al., 2024).

Hydrogen from alkaline electrolysis costs \$5.8/kg, less than the \$6.7/kg for a comparable PEM system (Hubert et al., 2024). In comparison, grey hydrogen from steam methane reforming costs \$2/kg (Jaeger and deBiasi, 2023), illustrating the large economic gap between traditional and green hydrogen synthesis. In the context of green ammonia plants, the Haber-Bosch process is often paired with alkaline electrolyzers to minimise hydrogen feed cost and maximise output, whereas plants centred on intermittent renewables favour PEM electrolysis for its rapid ramping capability (Hubert et al., 2024). As electrolyser costs fall, all these technologies are expected to play major roles in future green ammonia supply chains.

The ammonia synthesis reaction ($\text{N}_2 + 3\text{H}_2 \leftrightarrow 2\text{NH}_3$) is an exothermic reaction and involves a decrease in entropy. Per Le Chatelier's principle, the reaction is favoured by high pressure and low temperature. However, the slow kinetics of the reaction demand higher temperatures to be used in the Haber-Bosch process (Giddey et al., 2017). Due to a low single-pass conversion, typically around 10-15% per reactor (Rouwenhorst et al., 2021), a recycle loop is required to make the process economically feasible. Inert components in the feed, such as argon, do not react and accumulate within the loop. The build-up of these inert gases quenches the reaction, making a purge necessary to regulate their concentration in the reactor feed (Gaines, 1979).

To overcome equilibrium constraints and maintain conversion, ammonia synthesis is typically conducted in a multibed adiabatic system with interstage cooling. This reactor design prevents excessive temperature rise and keeps the reaction in favourable operating conditions (Azarhoosh et al., 2004). The standard kinetic model used for ammonia synthesis is the Temkin-Pyzheev equation, based on the assumption that dissociative adsorption is the rate determining step (Appl, 1999). Iron-based catalysts, such as wüstite, perform optimally at temperature ranges of 400°C-520°C (Tripodi et al., 2021) before deactivating rapidly at higher temperatures. The wüstite catalyst was found to retain 78% of its original reaction rate after exposure to a sintering treatment at 600°C (Pernicone et al., 2003). Recent studies denote that wüstite catalysts can

outperform magnetite-based catalysts by up to 70%, enabling low pressure operation (Liu et al., 2020).

Heat integration is a vital part of the synthesis loop, with the reactor effluent utilised to preheat the feed gas through a heat exchanger. A fixed UA is commonly used in steady-state simulations to reflect the physical limitations of the heat exchanger area and prevent equipment resizing. Maintaining reactor inlet temperatures at off-design loads requires either exchanger oversizing or bypass control strategies (Veltroni et al., 2025) and without sufficient heat recovery, the reaction may fall below its activation threshold, leading to a loss of conversion and reactor instability.

Upstream of the reactor, the purity and composition of the feed gases are equally critical. Nitrogen is typically sourced from cryogenic air separation, which delivers high purity, essential for preventing catalyst poisoning by oxygen. Alternatives such as pressure swing adsorption (PSA) offer greater flexibility but at the cost of purity and energy efficiency, rendering cryogenic systems the standard for large-scale plants (Rojas, 2024). The reactor feed must be compressed to around 150 bars, using multistage compressors with interstage cooling. Compressors are better suited for constant operating conditions rather than variable throughputs due to their narrow operating range. Compressors suffer from reduced isentropic efficiencies as the flow deviates from the design point compressors, and surge risk at low flows (Stewart, 2019).

While interest in flexible ammonia production has grown, the current ammonia synthesis loop is engineered to be inflexible and run uninterrupted at high capacity (Martinez, 2023). Since demand for green ammonia is rising, it is imperative for the conventional Haber-Bosch process to be redesigned for flexible operation to handle the intermittent hydrogen supply. A recent study showed that a load factor of 10% could be achieved while maintaining stability and efficiency (Patel et al., 2025). However, limitations imposed by compressor performance and heat exchanger sizing were not considered. These studies point to emerging flexibility but leave gaps in steady-state performance of the ammonia synthesis loop under significant throughput deviations. The loop's performance and constraints under partial load or overload scenarios, remain largely unquantified.

Having identified the gaps in the field of green ammonia, the following objectives of our study were decided:

- Build a full steady-state Aspen Plus model of an optimised three-bed reactor train with functional recycle and purge.
- Conduct a throughput analysis by observing the model's response across a wide range of throughputs and modelling the off-design compressor efficiency and fixed-UA heat exchanger.
- Ascertain the realistic operating range by executing a feasibility analysis and determine which equipment limit flexibility.

These objectives provide the insight needed to assess whether existing industrial infrastructure can support green ammonia production without radical redesign.

Methodology

Process Modelling Approach

A fully integrated Haber-Bosch process model was constructed using Aspen Plus V11 (Aspen Technology Inc, 2019), which was selected as it offers an extensive library of unit operation models and robust thermodynamics for gas liquid equilibria. Each subsystem was added sequentially onto a single flowsheet: air separation, hydrogen introduction, multistage compression, reactor train, heat integration, expansion and ammonia separation, and recycle, with the reactor train optimisation being conducted externally before integration.

Initially, the Peng-Robinson with Boston-Mathias modifications (PR-BM) property method was appointed as the global thermodynamic framework. PR-BM was recommended for high pressure, non-ideal gas mixtures and has been widely validated in ammonia synthesis modelling (Aspen Technology Inc, 2001).

Air Separation Unit

Once the method selection was complete, a cryogenic distillation flowsheet was developed on Aspen, designed to achieve 99.99% purity. Cryogenic distillation was preferred to pressure swing adsorption due to its large-scale, high-purity applications (Rojas, 2024).

The air inlet was passed through a splitter, followed by a high-pressure and low-pressure column and a system of coolers, similar to standard air separation units used industrially (Hüseyin Berat Erdöl, Fatma İrem Şahin and Nil Acaralı, 2024). Nitrogen purity and recovery were validated within Aspen against the specified targets (Rosbo et al., 2025), with high nitrogen purity desired, to reduce catalyst degradation due to oxygen impurity (Rossetti, 2020).

Multistage Compressor Train

Hydrogen was introduced as an external feed at 500,000 kg/day, matching renewable electrolytic production rates of Liquid Alkaline Electrolysers (James et al., 2023).

Subsequently, a multistage compressor train with interstage cooling compressed the gas to the reactor operating pressure. Each stage applied a 2.5 compression ratio (Baniamerian et al., 2024), with interstage cooling utilised to constrain discharge temperature below 177°C to protect compressor materials, seal components, and clearances (Stewart, 2019). The interstage coolers were modelled as heat exchangers, modifying cooling water flowrate to output the desired temperature of the stream, ensuring the discharge temperature was below 177°C.

Reactor Train and Kinetic Model

Following adjustment of the feed stream to its designated operating pressure and temperature, a three-bed reactor synthesis train was sequentially composed. Reactor volumes were increased from R1 to R3, emulating industrial practice where conversion rises as equilibrium becomes favourable at lower temperatures.

Table 1: Wüstite Catalyst Parameters inputted into the Aspen flowsheet to model the reactor train (Tripodi et al., 2018)

Parameter	N ₂	H ₂	NH ₃	Value
Stoichiometry	-0.5	-1.5	1	-
Rate Exponents				
Forward term (v_F)	1	2.25	-1.5	-
Reverse term (v_R)	0	-0.75	0.5	-
Kinetic Constant				
E_a (kcal mol ⁻¹)	-	-	-	45
k_0 (kmol s ⁻¹ kg _{cat} ⁻¹)	-	-	-	7.47×10 ⁸
Coefficients	A	B	C	D
Term 1 (K_{for})	-7.8	9218	-5.42	7.8×10 ⁻⁴
Term 2 (K_{rev})	2.88	0	0	0

The reaction kinetics were implemented using the Temkin-Pyzhev expression, calibrated to published wüstite catalyst parameters (Tripodi et al., 2018), illustrated in Table 1. The behaviour of the wüstite catalyst was modelled without adsorption terms, with simpler power-law rates, seen in Equation 1. K_a represents the equilibrium constant, $\frac{d\eta}{dt}$ represents the rate of consumption of the defective reactant, k is the kinetic constant of the direct reaction and a_i represents the activities of reactants and product. $\lambda(q)$ is a stoichiometric parameter set to 1 when the H₂:N₂ feeding ratio is 3. A reactor bed voidage of 0.33 and a catalyst particle density of 3.25 g cm⁻³ were adopted from literature values reported for industrial wüstite catalysts (Tripodi et al., 2018).

$$\frac{d\eta}{dt} = k \lambda(q) \left[(K_a)^2 a_{N_2} \left(\frac{a_{H_2}^{2.25}}{a_{NH_3}^{1.5}} \right) - \left(\frac{a_{NH_3}^{0.5}}{a_{H_2}^{0.5}} \right) \right] \quad (1)$$

After configuring the reactor kinetics, a separate model was developed to perform a sensitivity analysis on reactor length and inlet temperatures, with the aim of maximising ammonia conversion while avoiding excessive temperature rise. Reactor column diameters were constrained between 1.0 and 3.0 m, consistent with typical industrial reactor dimensions, while the length-to-diameter ratio (L/D) was restricted to 10 (Bittante et al., 2014). A temperature limit of 550 °C was also imposed due to catalyst deactivation constraints (Patel et al., 2025). The optimised lengths and reactor inlet temperatures were reintegrated into the main flowsheet.

Upon completion of the reactor train setup, a heat exchanger was implemented to recover energy internally and ensure sustainable practices in the design. Following common industrial Haber-Bosch configurations, the R1 effluent was used to preheat the R1 inlet, reducing external heating duty. Additional exchangers were implemented to adjust the inlet temperatures of each reactor. The R1 feed was preheated, and the R2 and R3 feeds were cooled using interstage coolers. The inlet temperature setpoints were reached by varying the flow of water for cooling and molten salt for heating. Molten salt was selected for its high thermal tolerance, thereby enabling efficient heating of the R1 inlet to the operating temperature.

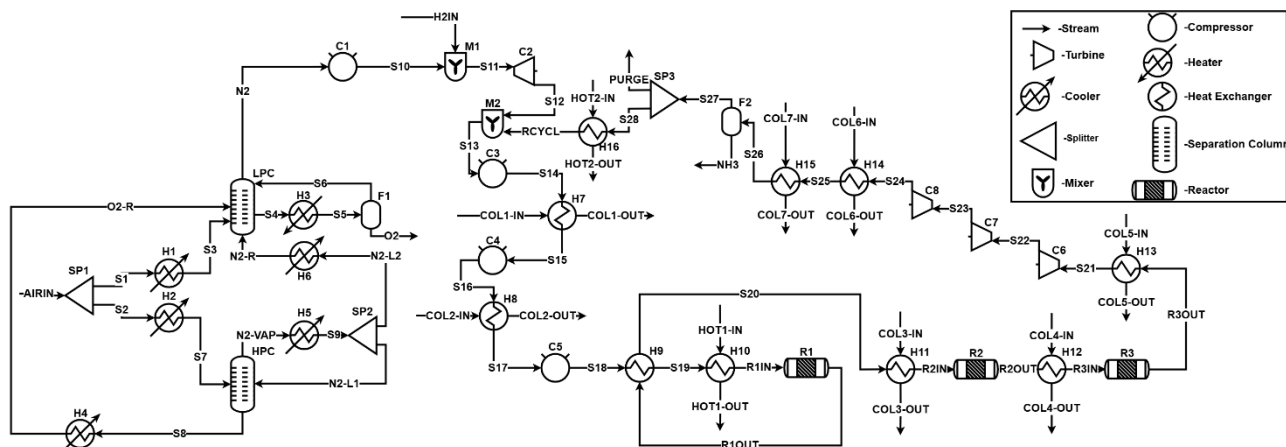


Figure 1: Process Flow Diagram (PFD) of the integrated ammonia synthesis plant developed on Aspen. The schematic details the primary unit operations, including the cryogenic Air Separation Unit (left), the multi-stage syngas compression train (C3–C5), the three-bed reactor synthesis loop (R1–R3), and the downstream ammonia separation section utilizing an expansion turbine train (C6–C8). A network of heat exchangers is incorporated around the process.

Ammonia Separation

Since the reactor effluent exited at high pressure and temperature, the stream was expanded and cooled downstream to reach the optimal temperature range for ammonia liquefaction and separation.

First, the reactor effluent was cooled below 177 °C, to prevent turbine performance deterioration (Stewart, 2019). Subsequently, a turbine train was incorporated, depressurising the effluent gas, whilst also recovering a small amount of work and lowering temperatures prior to condensation. Following the turbine train, heat exchangers with liquid nitrogen were utilised to liquify ammonia to cryogenic temperatures, employing a flash separator to yield high-purity ammonia.

Recycle Loop

To mitigate the low single-pass conversion characteristic of the Haber-Bosch Process, a recycle loop was introduced to a feed mixer upstream of the multistage turbine train. The design also integrated a purge stream at 5% within the recycle loop, maintaining impurities such as oxygen at a minimal level to curb catalyst degradation.

Since recycle compositions strongly influence reactor inlet conditions, an iterative manual adjustment strategy with an open recycle loop was employed to converge the reactor inlet and outlet properties. This process required reoptimising the reactor during each iteration, as the optimised parameters for reactor length and inlet temperature deviated with the changing stream composition.

Once the molar flows and component compositions of the inlet and outlet streams converged, the flowsheet for the steady-state simulation of the Haber-Bosch process was complete, illustrated in Figure 1. The optimised parameters for reactor length and inlet temperature in the final iteration were implemented in the steady-state flowsheet.

Throughput Analysis Overview

With the steady-state process model converged, a comprehensive throughput analysis was conducted. The purpose of the throughput analysis was to determine the operational range of the steady-state process model across varying hydrogen feed rates, while conforming to

equipment and kinetic constraints. This flexibility is essential for the green ammonia process, since the intermittent nature of renewable energy leads to fluctuating electrolyser output of hydrogen.

Throughput was varied by proportionally adjusting the hydrogen and air inlet streams in 5% increments, from 25% to 135% relative to the design throughput. For each throughput, the flowsheet was converged to an operationally viable steady state. Initial throughput simulations were carried out at fixed reactor inlet temperatures, where the process displayed limited flexibility. To introduce flexibility into the system, the reactor inlet temperatures were varied with throughput, since reactor inlet temperature is routinely used as a control parameter in ammonia synthesis reactors (Shamiri and Aliabadi, 2021). Flexibility was also introduced by modifying mass flows of cooling and heating fluids in the heat exchangers, maintaining the outlet temperature of the process stream at a fixed value. Allowing this degree of freedom in the model significantly broadened the feasible operation region, enabling stable operation across the full set of throughput changes.

To observe the impact of throughput variance on system performance, several key metrics were collated and evaluated. The analysis prioritised heat exchanger duty and compressor power to determine operational limits associated with varying throughputs. Concurrently, the reactor thermal sensitivity and conversion response to throughput changes were monitored to understand stability of the process.

Heat Exchanger Specification

The steady-state process model performed design calculations derived from temperature specifications of the stream outlet. Under off-design conditions, Aspen Plus automatically adjusts heat transfer area, altering the sizing of the heat exchangers, producing physically infeasible results.

Thus, to enforce realistic behaviour and maintain constant exchanger size, UA values from the 100% steady-state process model were extracted and fixed for the throughput analysis. The fixed UA approach guaranteed that changes in heat duty across the analysis were driven solely by flow and temperature differences rather than by dynamic geometry changes executed by

the simulation. Subsequently, the heat duties corresponding to each viable throughput for each heat exchanger were collated.

Off-Design Compressor Behaviour

Similar to the challenges observed in heat exchanger sizing, the compressor and turbine units were also found to be resized for every throughput variation by the simulation. To maintain realistic compressor behaviour, isentropic efficiencies were calculated externally for each compressor and turbine unit using Aungier's isentropic efficiency correlation for all throughput changes (Aungier, 2000). Aungier's equation is illustrated in Equation 2, where η is the isentropic efficiency, η_{max} is the peak efficiency, Q_{red} is the ratio of the flow rate at the varied throughput to the design flow rate and k_1 and k_2 are empirical curve-fitting constants.

$$\eta = \eta_{max} \left[1 - k_1 (Q_{red} - 1)^2 - k_2 (Q_{red} - 1)^3 \right] \quad (2)$$

The computed isentropic efficiencies inputted prior to resolving the flowsheet.

Feasibility Analysis Overview

The heat exchanger duties and compressor powers collated from the throughput analysis were processed in a MATLAB script that identifies the largest throughput interval in which every piece of equipment remains within its operable limit, based on chosen turndown ratios for each process equipment.

The script determined the design duty, P_{design} , of each process equipment as the maximum absolute duty observed across the throughput analysis. The turndown ratio, TR , of the equipment was applied, to calculate the minimum allowable duty, P_{min} , using Equation 3.

$$P_{min} = \frac{P_{design}}{TR} \quad (3)$$

For each throughput point, the script checked whether the required duty for all process equipment remained within P_{min} and P_{design} for each equipment simultaneously. The smallest throughput that satisfied this criterion for all the process equipment was identified as the minimum feasible operating throughput. The maximum feasible operating throughput was determined as the largest throughput produced by the Aspen model, therefore establishing the operating range. The limiting equipment were defined as the units that constrained the system's operational flexibility.

Results

In line with the project objectives, the following results validate the steady-state process flowsheet developed in Aspen Plus, before illustrating the effect of varying throughput on the process equipment and the subsequent feasibility analysis.

Reactor Behaviour at Design Throughput

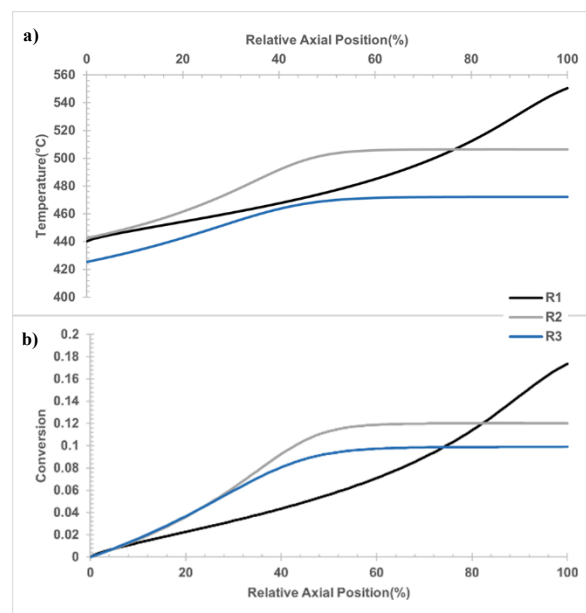


Figure 2: Axial profiles for reactors R1–R3 at design throughput. (a) Temperature distribution along the length of each reactor bed. (b) Conversion profiles highlighting the reaction progress across the three-bed reactor.

Upon the completion of the steady-state process flowsheet, the process validity was examined through an assessment of the reactor behaviour. Peak reactor temperatures of 549°C (R1), 550°C (R2) and 472°C (R3) were observed in the reactor bed, as seen in Figure 2a, preserving catalyst activity by operating within safe limits (Patel et al., 2025). Figure 2a also displays the characteristic temperature rise associated with the exothermic nature of the Haber-Bosch process. R1 reached the highest peak temperature, whereas R2 and R3 displayed diminishing temperature increases due to decreasing reactant availability.

The axial conversion profile shows the expected declining conversion from R1 to R3: the optimised single-pass conversions for each reactor were found to be 17.3% (R1), 12.0% (R2) and 9.9% (R3), visualised in Figure 2b. Consistent with the temperature trends, R1 provided the largest increase in conversion, while R2 and R3 delivered declining conversions as the reaction approaches equilibrium. A final single-pass conversion of 38.8% was achieved, falling within the expected industrial range for a standard Haber-Bosch process reactor train (Patel et al., 2025).

Effect of Throughput on Reactor Inlet Conditions

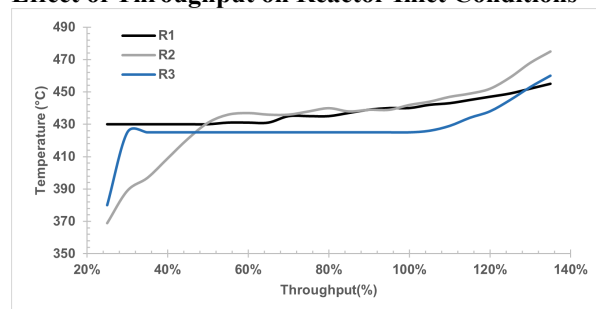


Figure 3: Graph showing reactor inlet temperature setpoints for R1–R3 across the throughput range, illustrating how inlet temperatures adjust under off-design conditions.

After validating the process model, throughput analysis was conducted by proportionally scaling hydrogen and air feed rates and resolving the Aspen flowsheet. The reactor inlet temperature was used as a control parameter to resolve the flowsheet, with the temperature response to throughput displayed in Figure 3.

R1 inlet temperature increased proportionally with throughput, rising incrementally from 430 °C to 455 °C. Conversely, R2 and R3 inlet temperatures displayed non-linearity: at low throughput, both R2 and R3 inlet temperatures fell substantially while at higher throughput, inlet temperatures rose sharply due to higher duty requirements.

Furthermore, the overall hydrogen conversion remained approximately constant at 91% across most converged throughput conditions; however, it fell substantially at low throughput, due to the collapse of the R2 and R3 inlet temperatures. Since the inlet hydrogen and air flows scale proportionally with throughput, the total ammonia production rate increased linearly, with higher throughput producing correspondingly greater quantities of ammonia.

Heat-Integration Behaviour

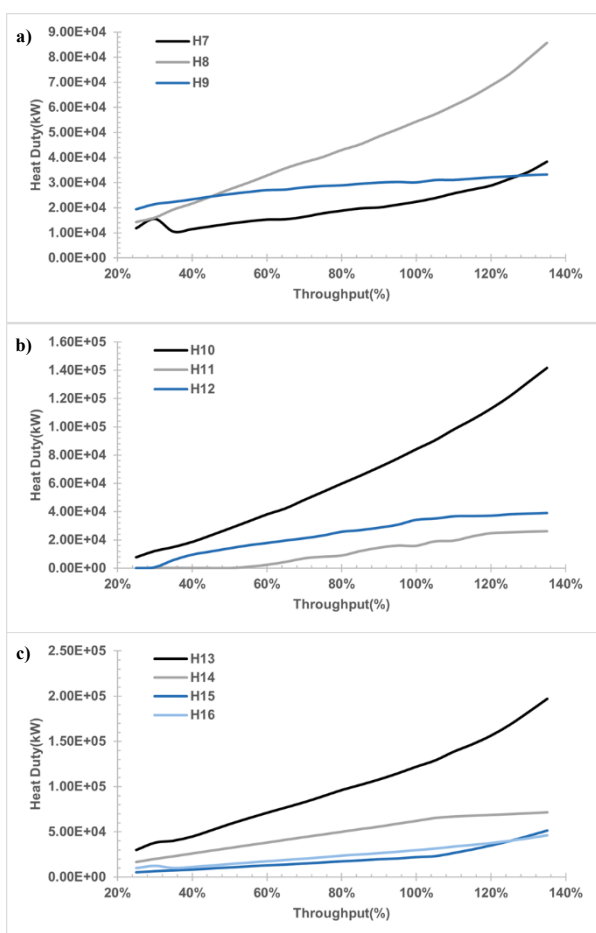


Figure 4: Variation of heat duty against throughput for process heat exchangers. (a) Compressor discharge cooling (H7, H8) to maintain temperatures below 175 °C, and feed preheating (H9) using R1 effluent. (b) Reactor loop thermal management (H10–H12), comprising additional feed preheating and interstage cooling between reactor beds. (c) Downstream processing units (H13–H16) responsible for cooling R3 effluent to –113 °C for ammonia separation and reheating the recycle gas.

At this stage, the UA values from the 100% steady-state process model were extracted and fixed for each heat exchanger, with Figure 4 displaying the heat duties of the heat exchanger network as throughput is varied. The heat-duty profiles exhibit clear non-linear behaviour, with certain exchangers, such as H10, showing strong sensitivity to throughput, while others, including H9, display comparatively weak sensitivity. The sensitivity of H10 can be explained by the R1 inlet temperature, which undergoes a sharp increase at high throughputs. In contrast, the duties for H11 and H12 fall towards zero at low throughputs, since inlet reactor temperatures for R2 and R3 decrease substantially at low throughputs, as seen in Figure 3. Therefore, the feed to R2 and R3 would require little to no cooling at low throughputs.

Compressor and Turbine Performance

Compressor and turbine shaft power were evaluated over the full throughput range. Figure 5 displays the exponential response of the compressor and turbine shaft power to throughput variance. The exponential response can be attributed to the isentropic efficiencies decreasing as the throughput varies further from the design, as per Aungier's correlation, seen in Equation 2.

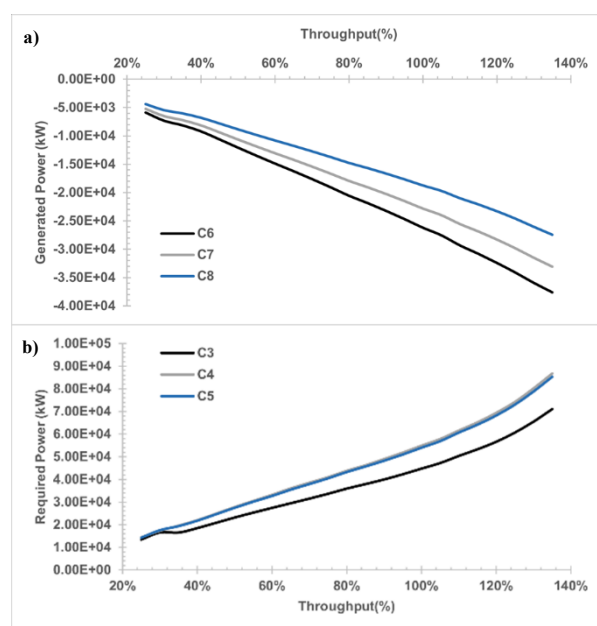


Figure 5: Variation of shaft work against throughput for the compression and expansion trains. (a) The three-stage feed compression train (C3, C4, C5) raising synthesis gas to 150 bars for reactor entry. (b) The expansion train (C6, C7, C8) depressurizing reactor effluent from 150 bar to 10 bar to enable ammonia separation.

Feasibility Analysis

After conducting throughput analysis and collating the heat duties as well as the compressor and turbine power, feasibility analysis was completed to identify the largest continuous operating range in which all major equipment remained within feasible limits. Table 2 summarises the minimum power requirements and throughput boundaries derived for process equipment. Heat exchangers exhibited a general ability to perform within feasible limits over all throughputs down to at least 25–35% capacity. A set of heat exchangers, namely H10, H11 and H15, imposed tighter constraints due to their high design duties, which resulted in comparatively

elevated minimum feasible duties. Among these, H11, with a turndown ratio of ten, imposed the most stringent constraint, restricting its minimum feasible throughput to 65%. H10 and H15 with higher duties and turndown ratios of four, were infeasible below 60% throughput.

The compressor train, with a minimum feasible throughput of 65%, was found to be the most restrictive item within the process flowsheet, hence determining the minimum feasible throughput for the process overall.

Thus, the feasible range for the process model was found to be between 65% and 135%. The lower bound on this range was found to be limited by compressor turndown ratios, while the upper bound was limited by model convergence, as the heat exchangers exceeded their design UA. This result establishes the feasible range for the green ammonia process with equipment turndown constraints. Hence, the compressor train is identified as the bottleneck in determining minimum feasible throughput, although the heat exchanger network also imposes a significant limit on turndown.

Table 2: Table showing the minimum operable duties, turndown ratios and feasible throughput limits for all major equipment in the process model, identifying the equipment that constrains the feasible operating range of the process.

Equip- ment	Specified Turndown	Design Power [kW]	Min Operable Power [kW]	Min Through- put [%]
H1	5	-8684	-1736.8	30
H2	5	-7787.8	-1557.6	30
H3	4	7143.5	1785.9	35
H4	5	-240.29	-48.059	30
H5	5	-3958.7	-791.74	30
H6	5	-221.67	-44.334	30
H7	4	38296	9574.1	25
H8	4	85651	21413	40
H9	4	33307	8326.8	25
H10	4	141590	35399	60
H11	10	26125	2612.5	65
H12	4	39203	9800.7	45
H13	4	197050	49263	45
H14	4	71528	17882	30
H15	4	51326	12831	60
H16	4	46228	11557	30
C1	2.5	7056.2	2822.5	55
C2	3	-9684.4	-3228.1	50
C3	2.5	71140	28456	65
C4	2.5	86733	34693	65
C5	2.5	85326	34130	65
C6	3	-37586	-12529	55
C7	3	-33074	-11025	55
C8	3	-27415	-9138.3	55

Convergence and Model Robustness

The steady-state model converged reliably across the central throughput range, without any warnings or errors. However, convergence closer to the extremes of the model's feasible region posed challenges, and required iteration adjustments, since the system is approaching physical feasibility limits.

Discussion

Reactor Behaviour at Design Throughput

The initial objective was to build a steady-state Aspen model with an optimised reactor train. The temperature and conversion profiles, shown in Figure 2, establish a baseline for the steady-state behaviour of the reactor train at the design throughput. The reactor behaviour aligns with prior models, indicating that the Temkin-Pyzhev kinetic model used provides an accurate representation of the reactor train's performance.

A key feature of the steady-state reactor profiles, shown in Figure 2, is the strongly front-loaded nature of the ammonia synthesis in the reactor train. The largest temperature rise and achieved conversion occur in R1, where reaction rates are highest due to the greater inlet concentrations of reactants. In R2 and R3, both temperature and conversion decline sequentially because reaction rates fall as reactant concentrations drop, and equilibrium is approached. The levelling observed in the later reactor beds reflects the diminishing driving force near equilibrium. This behaviour is typical in multi-bed ammonia converters where excess catalyst is used to sustain reaction rates and evenly distribute the heat of reaction.

With this baseline established, the model can progress to the next objective of assessing the effect of throughput variation on key performance metrics.

Reactor Behaviour at Off-Design Throughput

To accomplish throughput analysis, the Aspen flowsheet needed to be converged at each throughput. However, the model proved to be highly inflexible due to the equipment-sizing constraints imposed, such as the fixed UA values for the heat exchangers and the decreasing isentropic efficiencies for the compressors. To engineer flexibility, reactor inlet temperatures were adjusted manually and used as independent variables for each throughput. The reactor behaviour at off-design throughputs is assessed through analysing the change in the reactor inlet temperature setpoints.

The effect of varying throughputs on the required reactor inlet temperatures is depicted in Figure 3. While the inlet temperatures for the R1 reactor remain close to their design point value of 440°C, the temperatures for the R2 and R3 inlet streams vary more dramatically as throughput moves away from the design point. At low throughput, the fixed UA specification on the H9 heat exchanger leads to excessive cooling of the R1 effluent, as the reduced mass flow increases the effective temperature driving force across the exchanger. The absence of reheating between reactor beds results in the overcooled R1 effluent being passed onto R2 and R3, restricting both reactors to operate at substantially lower temperatures and resulting in severely limited conversion. At high throughput, the larger temperature rise across R1 increases the thermal load delivered downstream. The fixed heat-integration network becomes unable to compensate for this temperature rise, resulting in elevated temperatures for the R2 and R3 inlet streams. The trends for the inlet temperature requirements indicate that constraints within the heat-integration network play a significant role in limiting the steady-state flexibility of the process.

Heat Exchanger Behaviour

A clearer picture of these flexibility constraints emerges when examining how the heat exchangers respond to changes in throughput. The duty trends in Figure 4 show that only certain exchangers impose meaningful limits on operability, as their heat duties deviate significantly from the design point with changing throughput.

At low throughputs, the limitations of the heat-integration system become most evident around H9. Figure 4a illustrates that the heat duty of H9 remains relatively constant across the operating range, since H9 is unable to adjust the flow rate of one of its heating or cooling streams to target a specific outlet temperature. When throughputs fluctuate, the temperature driving force varies to offset this change, resulting in a relatively constant heat duty. The lack of an independent variable to specify an outlet temperature for the H9 heat exchanger leads to the R1 effluent to be excessively cooled at low throughputs, limiting the flexibility of the process. The cooling duties of the H11 and H12 interstage coolers for the reactor train fall to zero at low throughputs due to already low R2 and R3 inlet temperatures. The system is therefore limited by a lack of heating at low throughputs.

At high throughputs, the most significant constraints arise around the heat exchangers directly upstream and downstream of the reactor train. The H10 heat exchanger is responsible for heating the preheated stream from H9 up to the required inlet temperature for R1. The rise in required inlet temperatures for R1 as well as the reduced preheating provided by H9 at higher throughputs leads to the heat duty for H10 to rise exponentially. With a fixed UA constraint, H10 reaches its design capacity limiting its ability to raise the feed to the required reactor inlet temperature. A similar constraint occurs downstream, with the H13 heat exchanger responsible for cooling the effluent of the reactor train before it is expanded. As throughput increases, the R3 effluent becomes both hotter and higher in mass flow causing the cooling duty for H13 to rise exponentially. The exchanger's fixed UA becomes insufficient to accommodate this as it becomes undersized for high throughput operation. Together, both H10 and H13 establish an upper bound on the feasible operating range due to insufficient heating and cooling capacity.

Overall, these results show that the heat-integration network imposes both lower and upper bounds on feasible operation. At low throughputs, insufficient heating downstream of H9 prevents the reactor train from reaching the temperatures required for effective conversion, while at high throughputs the fixed UA constraints on H10 and H13 lead to operational limits being reached. These thermal constraints help define the model's steady-state flexibility and highlight the need for alternative heat-integration strategies if a wider operating range is to be achieved.

Compressor and Turbine Behaviour

Off-design compressor and turbine behaviour introduces another important constraint on throughput flexibility. The outlet pressure is specified in Aspen for each compressor and turbine as well as the isentropic

efficiency which decreases at off-design throughputs. The decrease in isentropic efficiency models the declining performance of the compressors and turbines when moving away from their design point. This explains the non-linear relationship of the power required with varying throughput displayed in Figure 5.

As throughput increases, the compressors and turbines are required to reach the same specified outlet pressure with higher flowrates and a decrease in isentropic efficiency. This leads to the power required for compression and the power generated from expansion to rise exponentially as compression and expansion become more energy intensive at higher throughputs.

The compressor and turbine performance at off-design throughputs places operational limits on the feasible throughput range of the model. These constraints act in parallel with the thermal constraints identified previously in the reactor and heat-integration network, collectively defining the steady-state operating range of the process.

Feasibility Analysis

The final objective was to determine the realistic operating range of the process and identify the equipment that limit flexibility. The feasibility analysis shows that the lower bound of operation occurs at 65% of design throughput, where the compressor train reaches its surge limit and becomes infeasible.

At higher throughputs, the feasible operating range is restricted due to the H10 and H13 heat exchangers reaching their operational limits. At these high throughputs, H10 and H13 are undersized and must exceed their fixed UA constraints in order to sufficiently heat and cool the reactor train inlet and outlet streams. Hence the upper operating limit is defined to be at 135% of design throughput.

Overall, the feasible steady-state operating range of the model is 65-135 % of the design throughput, bounded by compressor turndown and heat exchanger operational limits. To widen this range, flexibility in the heat-integration network can be introduced through the addition of interstage heaters in the reactor train. A bypass for the H9 heat exchanger can also be implemented to prevent the over cooling of the R1 effluent and keep R2 and R3 inlet streams at the required temperature at low throughputs. At high throughputs, increasing the number of coolers downstream of the reactor train would help distribute the high cooling duty required, preventing H13 from reaching its operational limit.

Furthermore, due to the low turndown ratio for compressors and turbines, it is vital to implement flexibility in this area when designing a green ammonia plant. One approach is to adopt parallel or multi-stage compressor configurations, where individual units can be engaged or disengaged in response to changes in throughput. Recycle compressors provide an alternative route to maintaining stable operation across a wider operating range however, this comes with a penalty in overall energy efficiency.

Conclusion

To conclude, this project aimed to evaluate whether a traditional Haber-Bosch process can be adapted for variable green ammonia production. This was conducted through building a steady-state Aspen Plus model of an optimised reactor train with recycle and purge, performing a throughput analysis with fixed-UA heat exchangers and off-design compressor behaviour, and conducting a feasibility analysis that identifies the realistic operating range in addition to the equipment that limits flexibility.

The core outcome of our project is a feasible steady-state operating range of 65-135% of design throughput. Two system-level phenomena dominate this range. At the lower bound of the operating range, the compressor train's limited turndown flexibility and surge behaviour prevent stable operation. At the higher bound, critical heat exchangers around the reactor train, namely H10 and H13, reach their duty limits, hence failing to provide the heating or cooling required by higher mass flows. Therefore, strong thermal coupling combined with limited compressor turndown set the practical feasibility limits of the process rather than reactor kinetics alone.

To translate these conclusions into a path towards wider and more resilient flexible operation, this report reveals three priority actions: improving the heat-integration strategy, reconfiguring compressors into recycle or multi-train compressors, and implementing more robust control of reactor inlet temperatures.

The heat-integration network in the model is the only process system that imposes both lower bound and higher bound limitations. At low throughputs, the heat exchangers overcool the R1 effluent and at high throughputs, the heat exchangers become undersized for required heating and cooling duties. Therefore, a heat-integration redesign can expand the feasible range substantially.

Beyond equipment improvement, the outcomes of the study provide two key takeaways. Firstly, a realistic flexibility benchmark is established for green ammonia plants considering physically constrained heat-exchanger sizing and realistic off-design compressor behaviour. Secondly, this study also creates a foundation for future dynamic modelling and plant redesign. This future work can assess the ramping between varying throughputs and identify process equipment that limit flexibility, enabling redesigns that improve flexibility in the Haber-Bosch process.

Furthermore, pilot-scale green ammonia plants would validate the modelling assumption of the Aungier off-design compressor efficiency used in this study. A clearer picture of practical integration issues such as start-up sequences would also be achieved, allowing for a more extensive study of the flexibility of the Haber-Bosch process. Studies examining alternative nitrogen production routes and different electrolyser technologies within an integrated system would also be valuable, as upstream choices influence both feed purity and operational flexibility. These effects directly interact with the constraints identified in this work and will shape the overall feasibility of green ammonia synthesis.

Anticipating future developments, the global transition to variable renewable energy will place even greater emphasis on understanding how green ammonia plants respond to frequent throughput variance. As the sector moves towards green hydrogen supply chains, future plants may require designs that are not only flexible in steady-state but inherently dynamic, capable of fast ramping, intermittent operation, and partial shutdown, while preserving catalyst health and equipment integrity. For future plants designed for variability from the start, rather than being retrofitted later, integrating short-term energy storage and control systems could offer promising pathways. In the context of this study, the insights developed serve as a stepping stone towards defining what a flexible Haber-Bosch system could look like.

In summary, this project demonstrates that conventional Haber-Bosch infrastructure can be adapted to green ammonia production, but only when the design explicitly accounts for heat-integration and compressor constraints. By prioritising adaptive heat-integration together with parallel compressors and tighter reactor temperature control, the Haber-Bosch loop can be redesigned from a rigid process into a more flexible process. Pursuing these lines of research will help translate the promise of green ammonia into industrial practice rather than an optimistic simulation.

Acknowledgements

The authors would like to thank Cameron Aldren for his continuous guidance and support throughout this study.

References

- Appl, M. (1999). Ammonia. Wiley-VCH.
- Aungier, R.H. (2000). Centrifugal compressors. American Society of Mechanical Engineers.
- Azarhoosh, M.J., Farivar, F. and Ale Ebrahim, H. (2014). Simulation and optimization of a horizontal ammonia synthesis reactor using genetic algorithm. *RSC Adv.*, 4(26), pp.13419–13429. doi:https://doi.org/10.1039/c3ra45410j.
- Baniamerian, Z., Garvey, S., Rouse, J., Cárdenas, B., Pottie, D.L., Barbour, E.R. and Bagdanavicius, A. (2024). How pressure affects costs of power conversion machinery in compressed air energy storage; Part I: Compressors and expanders. *Journal of Energy Storage*, 89, pp.111791–111791. doi:https://doi.org/10.1016/j.est.2024.111791.
- Bittante, A., García-Serna, J., Biasi, P., Sobron, F. and Salmi, T. (2014). Residence time and axial dispersion of liquids in trickle bed reactors at laboratory scale. *Chemical Engineering Journal*, 250, pp.99–111. doi:https://doi.org/10.1016/j.cej.2014.03.062.
- Engineering Toolbox (2019). Air - composition and molecular weight. [online] Engineeringtoolbox.com. Available at: https://www.engineeringtoolbox.com/air-composition-d_212.html.
- Gaines, L.D. (1979). Ammonia synthesis loop variables investigated by steady-state simulation. *Chemical Engineering Science*, 34(1), pp.37–50. doi:https://doi.org/10.1016/0009-2509(79)85176-3.
- Giddey, S., Badwal, S.P.S., Munnings, C. and Dolan, M. (2017). Ammonia as a Renewable Energy Transportation Media. *ACS Sustainable Chemistry & Engineering*, 5(11), pp.10231–10239. doi:https://doi.org/10.1021/acssuschemeng.7b02219.
- Guillermo, Ruiz-Gutiérrez, G., Viguri, J.R. and Galán, B. (2024). Flexible Green Ammonia Production Plants: Small-Scale Simulations Based on Energy Aspects. *Environments*, 11(4), pp.71–71. doi:https://doi.org/10.3390/environments11040071.
- Hubert, M., Esposito, A.M., Peterson, D., Miller, E. and Stanford, J. (2024). HYDROGEN SHOT: Water electrolysis technology assessment. [online] Available at: https://www.energy.gov/sites/default/files/2024-12/hydrogen-shot-water-electrolysis-technology-assessment.pdf.
- Hüseyin Berat Erdöl, Fatma İrem Şahin and Nil Acaralı (2024). Novel study on cryogenic distillation process and application by using CHEMCAD simulation. *ACS omega*, 9, pp.15165–15174. doi:https://doi.org/10.1021/acsomega.3c09490.
- Jaeger, H. and deBiasi, V. (2023). Electrolyzer advances reduce cost of green hydrogen. [online] *Gas Turbine World*. Available at: https://gasturbineworld.com/electrolyzer-advances/.
- James, B., Acevedo, Y., Prosser, J., Huya-Kouadio, J. and McNamara, K. (2023). Hydrogen production cost with alkaline electrolysis. [online] *Osti.gov*. Available at: https://www.osti.gov/servlets/purl/2203367
- KG Engineering Solutions (2023). Aspen plus: Simulation of cryogenic air separation. [online] *YouTube*. Available at: https://www.youtube.com/watch?v=wcH47kj2a4k.
- Liu, H., Han, W., Huo, C. and Cen, Y. (2020). Development and application of wüstite-based ammonia synthesis catalysts. *Catalysis Today*, 355, pp.110–127. doi:https://doi.org/10.1016/j.cattod.2019.10.031.
- Martínez, P.H. (2023). Grey-to-Green ammonia as a decarbonisation pathway: market assessment and techno-economic analysis. Thesis to obtain the Master of Science Degree in Energy Engineering and Management.
- Patel, R., Mehr, A.S., Valverde, J.I., Scheffknecht, G., Maier, J. and Zohourian, R. (2025). Studying the effect of dynamic operation conditions on green ammonia production synthesis loop. *Journal of Process Control*, 150, pp.103436–103436. doi:https://doi.org/10.1016/j.jprocont.2025.103436.
- Pernicone, N. (2003). Wustite as a new precursor of industrial ammonia synthesis catalysts. *Applied Catalysis A: General*, 251(1), pp.121–129. doi:https://doi.org/10.1016/s0926-860x(03)00313-2.
- Queiroz, S.S. and Costa, L. (2023). Green Hydrogen Power Supply: Challenges and Opportunities. 2023 8th IEEE Workshop on the Electronic Grid (eGRID), pp.1–6. doi:https://doi.org/10.1109/egrid58358.2023.10380851.
- Rojas, A.S. (2024). Production & Purification Ammonia, Nitrogen, And Green Hydrogen Production & Purification #1. [online] Berlin: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH. Available at: https://kh.aquaenergyexpo.com/wp-content/uploads/2024/02/Ammonia-Nitrogen-And-Green-Hydrogen-Production-Purification.pdf
- Rosbo, Joachim, Anker, D., Jensen, S., Skogestad, J., Jørgensen, J., Huusom, Jensen, D., Skogestad, S., Jørgensen, J. and Huusom, J. (2025). Highlights optimisation and robust control of a load-flexible haber-bosch ammonia synthesis loop optimisation and robust control of a load-flexible haber-bosch ammonia synthesis loop.
- Rossetti, I. (2020). Reactor design, modelling and process intensification for ammonia synthesis. *Sustainable Ammonia Production*, pp.17–48. doi:https://doi.org/10.1007/978-3-030-35106-9_2.
- Rouwenhorst, K.H.R., Van der Ham, A.G.J. and Lefferts, L. (2021). Beyond Haber-Bosch: The Renaissance of the Claude Process. *International Journal of Hydrogen Energy*, 46(41). doi:https://doi.org/10.1016/j.ijhydene.2021.04.014.
- Shamiri, A. and Aliabadi, N. (2021). Modeling and performance improvement of an industrial ammonia synthesis reactor. *Chemical Engineering Journal Advances*, [online] 8, p.100177. doi:https://doi.org/10.1016/j.cej.2021.100177.
- Smith, C., Hill, A.K. and Torrente-Murciano, L. (2020). Current and future role of Haber-Bosch ammonia in a carbon-free energy landscape. *Energy & Environmental Science*, [online] 13(2), pp.331–344. doi:https://doi.org/10.1039/C9EE02873K.
- Stewart, M. (2019). Surface production operations. Volume IV, pumps and compressors. Gulf Professional Publishing.
- The Royal Society (2020). Ammonia: Zero-carbon fertiliser, fuel and energy store. [online] Available at: https://royalsociety.org/-/media/policy/projects/green-ammonia/green-ammonia-policy-briefing.pdf.
- Tripodi, A., Compagnoni, M., Bahadori, E. and Rossetti, I. (2018). Process simulation of ammonia synthesis over optimized Ru/C catalyst and multibed Fe + Ru configurations. *Journal of Industrial and Engineering Chemistry*, 66, pp.176–186. doi:https://doi.org/10.1016/j.jiec.2018.05.027.
- Veltroni, V., Calabrese, M., Carcasci, C. and Pagliai, M. (2026). Techno-economic assessment of Power-to-Ammonia-to-Power for long-term energy storage. *Energy Conversion and Management*, 348, p.120621. doi:https://doi.org/10.1016/j.enconman.2025.120621.

Investigating the Impact of Electrode Material on the Synthesis of Carbon Nanotubes via Molten Salt Electrolysis

Rachel Rigby, Joseph Wright

Department of Chemical Engineering, Imperial College London, U.K.

Abstract Carbon nanotubes (CNTs) are valuable additives due to their exceptional mechanical, thermal and electrical properties, and are becoming increasingly relevant in energy storage applications. Molten salt electrolysis (MSE) offers a lower-energy alternative to conventional CNT synthesis while simultaneously using CO₂ as a feedstock, contributing towards carbon capture and utilisation (CCU) strategies aimed at net-zero emissions. This study investigates the production of CNTs using molten lithium carbonate at 770°C with a 2-hour electrolysis time preceded by a 15-minute activation period. Various electrode materials were tested, including iron and zinc-coated galvanised steel cathodes, and nickel and nickel-chromium anodes. The resulting products were assessed through their Faradaic efficiencies and crystallinities, the latter via Raman spectroscopy, to determine the best combination of materials. Other characterisation methods were employed including X-ray diffraction (XRD), simultaneous thermal analysis (STA) and scanning electron microscopy (SEM), to provide additional insights into the composition, morphology and overall quality of the CNTs. It was found that the combination of an iron cathode and nickel anode gave the best results overall. The addition of zinc to the cathode inhibited CNT formation and increased disorder. Anodic chromium content showed no clear correlation with crystallinity, indicating the potential suitability of chromium-containing alloys as corrosion-resistant electrodes in future work.

Keywords carbon nanotubes; molten salt electrolysis; electrode; electrochemistry; Faradaic efficiency; crystallinity

1.0 Introduction

Global energy demands continue to rise in response to rapid population growth, development and industrialisation, which comes at the cost of rising carbon dioxide emissions. Average atmospheric carbon dioxide levels are projected to reach 425.7 ppm in 2025. To achieve net-zero emissions by 2050 in alignment with the Paris Agreement's 2.0 °C temperature increase limit, a reduction of approximately 1.7 GtCO₂ per year is required [1][2]. Meeting these targets necessitates decarbonisation strategies that support the transition towards more sustainable energy production.

One solution is the widespread deployment of carbon capture and storage (CCS) technologies. However, its implementation is limited by high operational costs and the need for complex transport infrastructure especially between inland capture sites and offshore storage [3]. As a result, carbon capture and utilisation (CCU) has gained interest as a complementary approach, as it converts carbon dioxide into valuable materials such as carbon nanotubes (CNTs) [4] without the need for transport, therefore offsetting part of the cost associated with carbon capture technology.

A major challenge for CCU is the substantial amount of energy typically required to convert carbon dioxide [4], which undermines the economic and environmental benefits if not powered by low-carbon electricity. However, electrochemical production of multi-walled carbon nanotubes (MWCNTs) using carbon dioxide as a feedstock offers a promising exception.

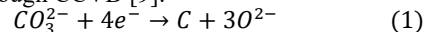
MWCNTs are a valuable additive to many materials to improve their chemical, electrical and mechanical performance, with applications such as conductive fillers in automotive plastics, reinforcement agents in polymers and composites, and potential replacements for copper in microelectronic interconnects [5]. Additionally, they can be added to lithium-ion battery electrodes, and rising energy storage markets could further increase demand for CNTs [6]. This is reflected through their market value and projected growth, with a sale price of approximately \$110 per kg [7], and a demand that is projected to grow at a global compound annual growth rate of 8.98% [6].

Currently, catalytic chemical vapour deposition (CCVD) is the most popular industrial method for synthesising MWCNTs due to its relatively lower cost and controllability compared to alternative methods [8]. However, it is still highly energy-intensive at 400 kWh/kg_{CNT} [9]. In contrast, using molten salt electrolysis (MSE) to produce CNTs has the potential to reduce energy usage to just 54 kWh/kg_{CNT} [9] while utilising CO₂ as a feedstock as an additional benefit. Further research in field is imperative to understand how process parameters influence the quality and yield of the product, to aid with scale-up efforts.

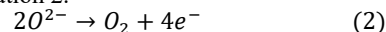
2.0 Background

Molten salt electrolysis (MSE) has emerged as a promising method to manufacture carbon nanotubes (CNTs). It works by submerging two electrodes in a molten salt and applying a current to drive electrochemical reactions.

At the cathode, carbonate ions are reduced to solid carbon and oxide ions, as demonstrated in Equation 1. The carbon attaches to nucleation sites where structures grow in a sp² hybridised form, resulting in strong hexagonal CNT structures that exhibit excellent mechanical, thermal and electrical properties. Presently, only MWCNTs have been produced using this method, although there is a possibility to produce single-walled carbon nanotubes (SWCNTs) by investigating the same catalysts that are used to produce SWCNTs through CCVD [9].

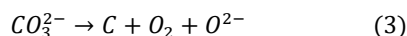


The anode's primary function is to promote the oxidation of the oxide ions, O²⁻, that are formed at the cathode, resulting in the formation of oxygen [10]. This is shown by Equation 2.

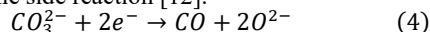


However, more recently researchers have hypothesised that small amounts of the anode material can also react with oxide ions and migrate to the cathode [11]. By choosing an appropriate anode material, this phenomenon can be beneficial as it can transfer metal ions that are effective CNT nucleation sites to the cathode.

Overall, the desired reaction during MSE is for carbonate ions to be converted to form solid carbon nanotubes and oxygen gas, through the following:



In addition to the production of carbon at the cathode, there is a side reaction where carbonate ions are reduced to carbon monoxide (Equation 4) [9]. While this reaction is not dominant at 770 °C, higher operating temperatures will favour the side reaction [12].

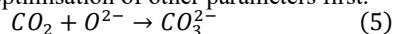


Extensive work has been performed to optimise parameters based on reaction efficiency and CNT quality. High temperatures must be employed to ensure the electrolyte is fully melted, and to increase its fluidity, thereby reducing resistance [13]. However, excessively high temperatures may promote the unwanted formation of carbon monoxide. Lithium carbonate has emerged as one of the leading electrolytes as it is cheaper and more thermally stable than the alternatives [14]. Its melting temperature of 723 °C constrains the operating temperature. Longer electrolysis times have been shown to produce more crystalline, tubular structures [15].

Another crucial consideration for producing high-quality CNTs is the choice of cathode material, as it must have an unfilled d^{10} orbital to be able to form a dative bond with carbon [16]. Transition metals such as iron and nickel have emerged as the best candidates [10]. However, there is a risk of oxide layer formation if pure metal is used. Alloys like galvanised steel, with a zinc outer layer can be used to combat this. It is interesting to note that zinc's melting point is 419.5 °C [17], which is significantly lower than the furnace operating temperature of 770 °C in Imperial College London's lab set-up. This may mean that the zinc in the galvanised steel melts and forms an iron-zinc ferrite phase during MSE [18], impacting the electrochemical reactions.

Iron, nickel and even chromium are also good materials for the anode as, similarly to the cathode, they can become effective nucleation agents for CNTs through transition metal deposition on the cathode [10][19]. Additionally, implementing an activation period into the synthesis is important for ensuring transition metal deposition can occur in the first place. Literature suggests that low current densities such as 5 to 10 mA/cm², which result in low potentials, are most suitable [20] as they allow selective transition metal deposition on the cathode, while preventing carbonate ions from being prematurely reduced to carbon.

A carbon dioxide stream is necessary for regenerating the carbonate ions as shown in Equation 5, to ensure continuous production of CNTs is achieved and all other materials are not consumed. Other research groups have successfully attempted this by bubbling CO₂ gas into the molten salt [21]. However, this has not yet been achieved at Imperial College London due to prioritising the optimisation of other parameters first.



In this study, various electrode materials are tested and compared based on Faradaic efficiency and crystallinity to determine the best combination of materials. Galvanised steel and pure iron are used as cathodes to examine how the presence of zinc impacts the product, while nickel-chromium alloys and pure nickel are used as anodes, both to confirm that transition metal deposition occurs and whether chromium offers additional corrosion-resistance and nucleation benefits.

Other parameters such as temperature, current density and synthesis time have been previously optimised through literature review and experimentation on the specific set-up in the Imperial College London lab. The specific values of each can be found in the methods section.

3.0 Methods

Electrolysis

Electrode Fabrication

Electrodes were prepared using wires of different materials (see Table 1). Each wire was threaded through an alumina support tube, and the ends were shaped into uniform spirals with surface areas of 3.14 cm².

Table 1: Electrode materials and their respective experiments

Exp.	Cathode Material	Anode Material
1	Iron (pure)	Nichrome
2	Iron (pure)	Inconel 625
3	Galvanised Steel	Inconel 625
4	Galvanised Steel	Nichrome
5	Galvanised Steel	Nickel (pure)
6	Iron (pure)	Nickel (pure)

Compositions of each material were obtained from X-Ray fluorescence (XRF) analysis and are presented in Table 2 below.

Table 2: XRF compositions in wt% for each material. Only elements above 1 wt% are included

Electrode	Fe	Zn	Ni	Cr	Mo	Nb
Iron	99.8	-	-	-	-	-
GS	56	44	-	-	-	-
Nickel	-	-	99.8	-	-	-
Nichrome	1	-	77	21	-	-
Inconel 625	-	-	65	22	8	3

Reactor Assembly

A cylindrical alumina crucible (internal dimensions $L = 311$ mm, $D = 26$ mm) was mounted vertically on a clamp stand, and 30.00 ± 0.08 g of lithium carbonate was added in a fume cupboard. The cathode and anode were placed in the crucible and secured using a series of flanges and O-rings so they would be submerged within the molten electrolyte but would not contact each other or the crucible wall during electrolysis. Stabilising attachments were added to secure the crucible in the furnace. The set-up is shown in Figure 1.

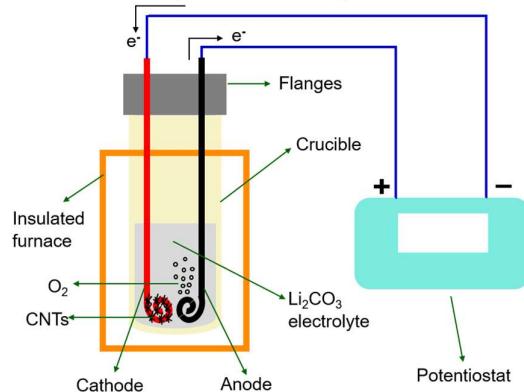


Figure 1: Schematic of the molten salt electrolysis set-up for CNT synthesis, showing the furnace, crucible and potentiostat

Electrolysis Procedure

The assembled reactor was heated in a Carbolite Gero furnace at 4 °C min⁻¹ to 770 °C, followed by a 30 min settling period. Electrodes were connected to the Autolab PGSTAT30 potentiostat and electrolysis was controlled galvanostatically (i.e. constant current was applied) using Nova 2.1 software with an activation period of 15 minutes at 0.0157 A (5 mA/cm²) followed by a 2-hour electrolysis period at 0.628 A (200 mA/cm²). Activation and electrolysis data were recorded and exported for analysis to gain insight into how the potential difference varies during electrolysis.

Post-processing

After cooling overnight, the crucible and cathode were washed in ~6 % HCl which was replaced hourly, to react with any excess lithium carbonate. CNTs were detached from the cathode by gentle mechanical agitation, vacuum filtered and dried overnight in an 80 °C oven. The dried product was weighed to measure yield and ground into a powder for subsequent analysis.

Faradaic Efficiency

To determine the efficiency and selectivity of CNT synthesis for each experiment and thus determine the performance of each electrode pair, their respective Faradaic efficiencies were calculated using Equation 6.

$$\text{Faradaic Efficiency, } \Phi = \frac{\text{Mass Produced}}{\text{Theoretical Mass}} * 100\% \quad (6)$$

Mass produced is merely the weight of dried product obtained, and the theoretical mass expected was calculated from known operational parameters and constants with the following equations:

$$\text{Theoretical Mass} = \frac{A_{r,C} * Q}{v_{e-} * F} \quad (7)$$

with $A_{r,C}$ representing molar mass, v_{e-} representing the electron stoichiometry, F representing Faraday's constant and Q representing charge, as given by:

$$Q = J * A_e * t_s \quad (8)$$

where J represents current density, A_e represents surface area of the electrode and t_s represents electrolysis time, excluding the activation period. Values for each parameter can be found in Table 3.

Table 3: Parameter values used in Faradaic efficiency calculations

Parameter	Parameter Value	Parameter Unit
$A_{r,C}$	12	g/mol
v_{e-}	4	-
F	96,485	C/mol
J	200	mA/cm ²
A_e	3.14	cm ²
t_s	7200	s

Economic Analysis

Using the Faradaic efficiency and the cell potential difference provided by the electrolysis graphs, an estimation for electrical energy consumption was calculated using Equation 9:

$$w_{CNT} = \frac{4F}{\Phi_{CNT}} * \frac{U_{cell}}{3600 * A_{r,C}} \quad (9)$$

where w_{CNT} represents the specific electrical energy consumption in kWh/kg_{CNT} and U_{cell} represents the cell

potential difference in V. From this, the electrical cost of producing CNTs was found by multiplying w_{CNT} by the unit cost of electricity per kWh, reported as 30 p/kWh at Imperial College London in May 2024 [22].

Analytical Testing

Raman Spectroscopy

Raman spectroscopy was used to analyse crystallinity. Samples from all experiments were placed in a Bruker Senterra II instrument and analysed over a wavenumber range of 50 to 4260 cm⁻¹ at 20× magnification using a 532 nm laser. To prevent burning and thermal degradation, a low laser power (2.5 mW) and 30 s integration time were employed. For each sample, data was collected over 4 points and averaged, discounting any anomalies to better account for any inconsistencies. An Asymmetric Least Squares (AsLS) method was implemented [23] to baseline correct spectroscopic data and ensure consistent baselines across samples for fairer, more accurate comparison.

Using the spectra produced, the D, G and 2D peaks were identified (1340 cm⁻¹, 1580 cm⁻¹ and 2690 cm⁻¹ respectively) and their intensities were used to calculate I_D/I_G and I_{2D}/I_G ratios, to assess the degree of disorder and interlayer structure of the CNTs.

X-ray Diffraction

X-ray diffraction (XRD) was used to identify the presence of CNTs and other contaminants. Samples were analysed in a Malvern PANalytical Aeris Diffractometer over a 2θ range of 10 - 85°. The instrument operated at 40 kW and 15 mA with a Cu-Kα radiation wavelength of 1.54 Å (0.154 nm). The data was then baseline corrected with an AsLS method to allow for more accurate peak identification [23]. Peaks were compared against known diffraction peaks for elements and compounds suspected to be found in the samples.

Thermal Analysis

Simultaneous Thermal Analysis (STA) was performed using a Netzsch STA 449 machine, allowing data to be collected for both mass change over time from thermogravimetric analysis (TGA), and heat flow behaviour from differential calorimetric analysis (DSC). Measurements were taken up to 1000 °C in an air environment, with a heating rate of 10 K/min.

SEM Analysis

Scanning electron microscope (SEM) images were produced using the Zeiss SEM machine at magnifications between 45k and 105k×. This gave insights into the CNT morphology and surface characteristics of each sample.

4.0 Results and Discussion

Activation and Synthesis

During activation and synthesis, the cell potential differences across the electrodes were measured. Potential is proportional to the energy required to drive electrochemical reactions and overcome ohmic and kinetic barriers. A higher potential indicates more difficulty or resistance within the system as current density remains constant, while lower potentials mean that the pathway to produce CNTs is more efficient.

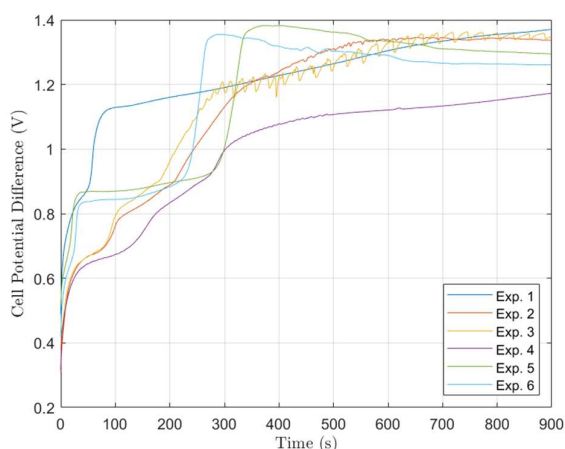


Figure 2: Graph of the activation period for all experiments performed at 5 mA/cm², showing how potential difference varied over the 15-minute timespan

To achieve the 1 V activation period necessary for promoting transition metal deposition, as discussed in the background section, a low current density was initially implemented.

Figure 2 shows that all samples exhibit an initial rapid rise in potential corresponding to the current density step-up at $t = 0$ s. The subsequent graph shapes support the idea that the activation period is linked to nickel and chromium activity, and therefore anode composition. Experiments 5 and 6, which were both synthesised using a pure nickel anode, have very similar shapes, characterised by an initial plateau at around 0.85 V, followed by a second plateau at around 1.3 V. The first plateau may be associated with the reduction of cathodic iron to form active nucleation sites at the surface of the cathode, while the second plateau may represent the nickel deposition that occurs at low current densities, as previously discussed in the background section. The lack of other ions such as chromium means the system is mostly steady, reflected by smooth curves.

Experiments 1, 2, 3 and 4 have irregular, wavy patterns that are distinctly different from the smooth plateaus in 5 and 6. This is likely due to the presence of chromium in the anode, which provides alternative reaction pathways. Instead of a straightforward nickel reduction, there is the possibility for chromium reduction, and even a reaction between chromium and nickel or the electrolyte. Another explanation is that chromium can form a passivating oxide layer at the anode surface [24] and while it protects the anode against corrosion, it may also inhibit nickel ions from consistently migrating to the cathode to be reduced. Experiment 3 shows the most fluctuation in potential, possibly due to both electrodes being metal alloys, so the additional zinc and chromium introduce further complexity into the system.

As shown in Figure 3, the step-up in current density to 200 mA/cm² at $t = 0$ s results in a rapid increase in potential as the system adjusts to the higher current, with all samples reaching a peak between 100 and 500 seconds. This is likely to represent the maximum initial energy required for the first carbonate ions to be reduced to carbon. Following this initial peak, the curves gradually decrease over time. This could be attributed to the formation of CNTs, which increase the surface area over which growth can occur and act as a strong conductor due to the delocalised π electrons from sp^2 hybridisation. Both aspects increase reaction

efficiency, therefore reducing the potential required to sustain the current density and reaction rate.

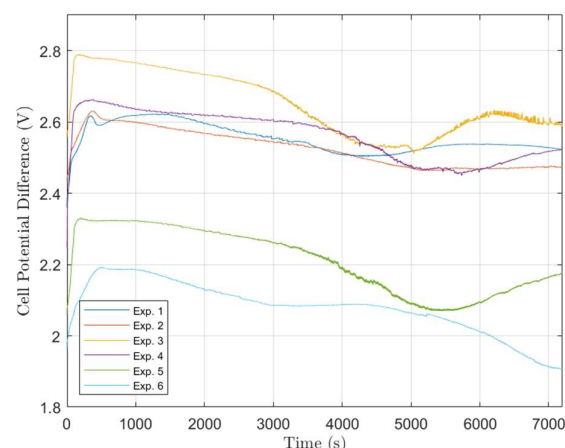


Figure 3: Graph of the synthesis period for all experiments performed at 200 mA/cm², showing how potential difference varied over the 2-hour timespan

Experiments 5 and 6 had the lowest potential throughout, peaking at 2.32 and 2.2 V respectively. This implies that their reaction pathways had the least resistance, leading to more efficient energy usage. Both experiments used a pure nickel anode, suggesting the anode reaction is more efficient without the presence of chromium. Although literature suggests that chromium acts as an effective nucleation agent for CNT growth [19], our results support the conclusion that pure nickel is still the better option at the anode.

Experiments 3, 4 and 5 were synthesised using a galvanised steel cathode and their potentials were higher than the corresponding iron experiments with the same anode. Experiments 3 and 4 in particular exhibited the highest potentials throughout, peaking at 2.8 V and 2.65 V, implying their reaction pathways had the highest resistance. Zinc is not reported as a suitable catalyst for CNT production due to its filled d^{10} orbital not being available to form a dative bond with carbon [16], so its presence in galvanised steel has no direct benefit for CNT growth. As discussed in the background, zinc's low melting point means that liquid zinc and/or an iron-zinc ferrite phase will be present in the reactor [18]. This may reduce the availability of iron for carbon nucleation, leading to an increase in resistance which explains the high potentials observed. However, while this suggests less efficient growth, the presence of zinc may still influence the crystallinity and morphology of the samples in a beneficial manner, so further characterisation is necessary.

It is also observed that the experiments that used an iron cathode, namely 1, 2 and 6, reached their maximum peaks slower than those produced using galvanised steel. One possible explanation may be that during the activation period, iron oxide forms in the absence of zinc which delays the initial reduction of carbonate ions as the nucleation sites are physically blocked. While this may imply that overall CNT growth is limited as it occurs over a shorter time, this does not appear to have occurred as Table 4 shows that these experiments did not have lower Faradaic efficiencies than the galvanised steel experiments.

Faradaic Efficiency

Table 4: Measured masses and calculated Faradaic efficiencies for each experiment

Experiment	Measured Mass (mg)	Faradaic Efficiency, Φ_{CNT} (%)
1	67.5	48.0
2	101.5	72.2
3	55	39.1
4	77.5	55.1
5	80	56.9
6	81	57.6

The Faradaic efficiencies found for each experiment are listed in Table 4. Generally, the calculated efficiency values are quite low, as the masses of CNTs obtained were significantly lower than the expected 140.6 mg. One possible explanation for this is due to the electrochemical side-reaction of the carbonate ions, forming carbon monoxide at the cathode. At the 770 °C operating temperature, this side-reaction is known to occur to some extent [9], and so a reduced proportion of the current contributed towards the formation of CNTs on the cathode, which resulted in lower product masses. Another possible explanation for the low product masses is due to loss between the crucible and the dried weighing stage. Of particular note, after vacuum filtration for all samples, a black solid residue was observed upon the filter paper, which indicated the loss of some CNT material. The error associated with this loss is hard to quantify given that the methodology followed did not involve weighing the filter paper before or after filtration. However, it is likely to be quite substantial given the known tendency of CNTs to bond closely to cellulose fibres [25], like those that make up filter paper.

The experiments that used pure iron as the cathode (Experiments 1, 2 and 6) instead of galvanised steel (Experiments 3-5) appeared to have higher Faradaic efficiencies, corresponding to higher yields. This could be due to the zinc coating within the galvanised steel hindering the availability of iron's nucleation sites for CNT formation as mentioned previously. Experiment 2 gave an unusually high measured mass of product when compared to the other experimental runs, which directly resulted in its higher Faradaic efficiency. This could be attributed to unexpected contaminants within the weighed sample. Notably during synthesis, the cathode adhered to the side of the alumina crucible. Removal could have inadvertently contaminated the sample, increasing its mass with non-carbon material.

Economic Analysis

Using average cell potentials across the 2-hour electrolysis period, the energy usage per kilogram was calculated with Equation 9. As shown in Table 5, they align with literature estimations of 54 kWh/kg_{CNT} [9]. Using an electricity cost of 30 p/kWh [22], electrical costs were also obtained. Experiment 2 had the lowest cost at 9.37 £/kg_{CNT}, although this may have been caused by contamination, so the best overall cost was achieved with Experiment 6 at 9.68 £/kg_{CNT}. Similar calculations show that using the same electricity cost, a 400 kWh/kg CCVD process [9] would have a much higher value of 120 £/kg_{CNT}, indicating our results support the usage of MSE over CCVD.

Table 5: Electrical energy usage and cost associated with each experiment, calculated using cell potentials and Faradaic efficiencies and averaged out on a per kilogram of CNT basis

Experiment	$U_{cell,av}$ (V)	w_{CNT} (kWh/kg _{CNT})	$Cost_{CNT}$ (£/kg _{CNT})
1	2.55	47.52	14.26
2	2.53	31.25	9.37
3	2.65	60.55	18.17
4	2.56	41.54	12.46
5	2.21	34.70	10.41
6	2.08	32.26	9.68

Raman Spectroscopy

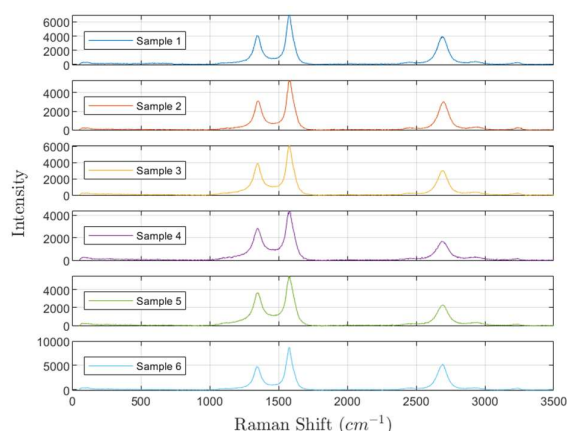


Figure 4: The obtained Raman spectra for all samples tested, showing the measured intensities for Raman shift between 50 and 3500 cm⁻¹. Samples represent their respective experimental runs as described in Table 1

Using the Raman spectra in Figure 4, valuable data regarding the degree of disorder and layering can be obtained. The D band, representing defects in CNTs [26] occurs at 1340 cm⁻¹, while the G band, representing the presence of sp² carbon networks (i.e. CNTs) [26] occurs at 1580 cm⁻¹. Additionally, there is a third peak known as the 2D band at 2690 cm⁻¹, which is related to the stacking order within CNTs [26]. Intensities at each of these bands can be used to calculate I_D/I_G ratio, with a lower value corresponding to less disorder and higher crystallinity. Additionally, the I_{2D}/I_G ratio can be calculated to assess how well-structured the MWCNT layers are [27]. This ratio allows for comparison between samples to identify those with fewer interlayer defects, as well as differentiating between SWCNTs and MWCNTs.

Table 6: I_D/I_G and I_{2D}/I_G ratios calculated from the Raman spectroscopy peaks for all samples

Sample	I_D/I_G ratio	I_{2D}/I_G ratio
1	0.59	0.56
2	0.58	0.57
3	0.64	0.50
4	0.65	0.39
5	0.66	0.42
6	0.55	0.61

According to the literature, MWCNTs produced using the conventional CCVD method, followed by chemical and heat treatment, exhibit I_D/I_G ratios between 0.2-0.9 [28]. The data in Table 6 indicates that the samples from all our experiments are sufficiently crystalline to be comparable with commercial CNTs. However, the lowest ratio (0.55) lies in the upper half of the reported range, indicating some defects are present. One possible source of these defects could be from the nickel deposition [10]. While this was a beneficial phenomenon for CNT growth and yield as discussed previously, excessive nickel on the cathode may alter the CNT diameter and uniformity.

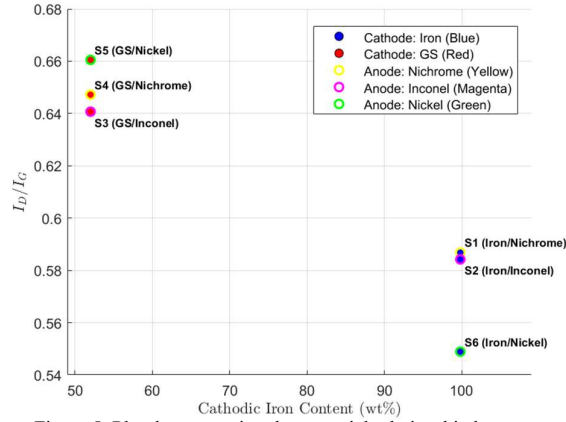


Figure 5: Plot demonstrating the potential relationship between cathodic iron content during synthesis and the resulting I_D/I_G ratio found from Raman spectroscopy, where a lower ratio indicates less defective CNTs

Figure 5 clearly shows how pure iron outperforms the galvanised steel cathode, suggesting that the presence of zinc is causing defects. As previously discussed in the background section, the zinc could melt or form a ferrite structure with iron, hindering access to growth sites and causing inconsistent growth.

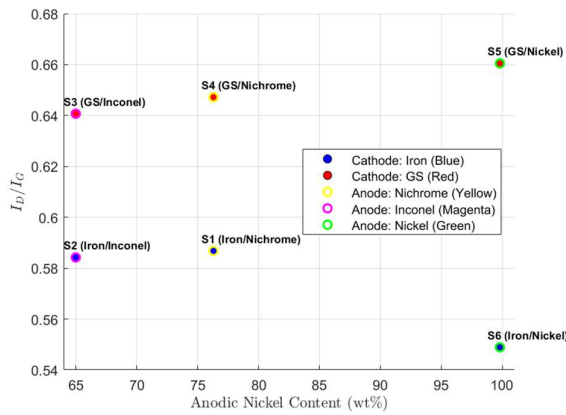


Figure 6: Plot demonstrating how the I_D/I_G ratio found from Raman spectroscopy does not appear to be related to the anodic nickel content during synthesis

Figure 6 shows that there is no clear trend between nickel composition in the anode and the degree of disorder. This means that chromium does not appear to cause significant defects and could be used in future experiments instead of zinc as a way to prevent iron oxidation.

All of the samples' I_{2D}/I_G ratios fall between 0.38 and 0.61 as shown in Table 6. Typical MWCNTs are reported to exhibit values less than 1 and SWCNTs show much higher values of above 2 [29]. This suggests that

all samples are multi-walled as expected. In MWCNTs, a higher I_{2D}/I_G ratio also reflects better ordering between layers. Generally, samples with a higher I_{2D}/I_G ratio such as the samples from Experiments 2 and 6 also have a lower I_D/I_G ratio, implying fewer structural defects overall.

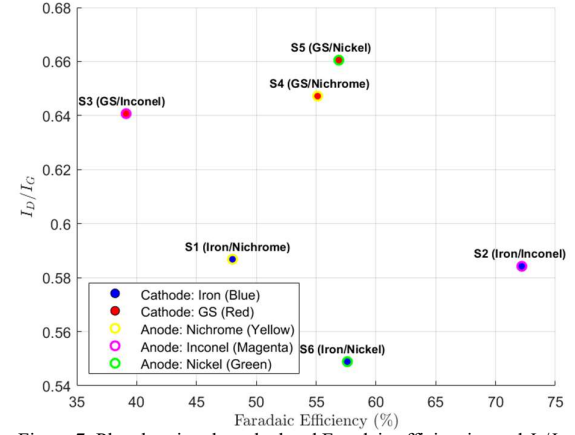


Figure 7: Plot showing the calculated Faradaic efficiencies and I_D/I_G ratios for all experiments conducted. The most successful experiments, with higher yields and less defects are found in the bottom-right

To identify the best sample, Faradaic efficiency was plotted against the I_D/I_G ratio in Figure 7. The optimal region is towards the bottom right of the graph, where Faradaic efficiency is highest while I_D/I_G ratio is lowest. Excluding Experiment 2 which as previously mentioned had an abnormally large Faradaic efficiency, Experiment 6 clearly demonstrated the largest, most crystalline yield. Furthermore, its sample has the highest I_{2D}/I_G ratio indicating that its interlayer structure is the most ordered.

XRD Analysis

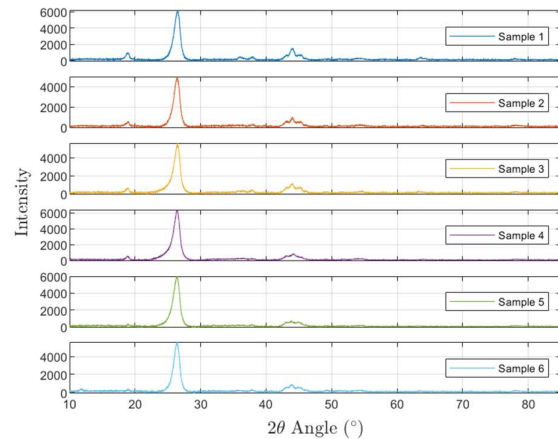


Figure 8: The obtained XRD spectra for all samples tested, showing the measured intensities for angles between 10 and 85°. Samples represent their respective experimental runs as described in Table 1

As shown in Figure 8, all samples exhibit the three most intense characteristic diffraction peaks for graphitic carbon [30]. The first is at 26°, corresponding to the (002) plane parallel to the graphene layers, and is very sharp which indicates consistent interlayer spacing and high crystallinity. The second is at 44°, and while it is much wider, this may be due to overlap with other diffraction peaks around 43-45°, rather than the presence of amorphous carbon. Additionally, even

crystalline graphitic carbon graphs in literature generally have a wider peak [19]. There is also a third, very small peak at 54.5°, corresponding to the (004) plane. This further supports that crystalline carbon is present as this peak is absent for amorphous carbon.

Samples 1 to 4 exhibit peaks at 18°, 36° and 44°, which align with lithium chromium dioxide's (LiCrO_2) diffraction peaks [15]. Since they were all synthesised using either nichrome or Inconel 625 anodes, which both contain chromium, it makes sense that the chromium could have reacted with the electrolyte and oxidised to form LiCrO_2 . Samples 5 and 6 were synthesised using chromium-free electrodes and show much smaller signals in these regions, which could be due to noise or contamination.

Diffraction peaks are also observed in all samples, at 52°, 64° and 78°. They are very small and may be attributed to metallic impurities, although it is difficult to identify specific elements without use of specialised analysis software.

To obtain qualitative composition results, Rietveld analysis is required which is too complex for samples that contain so many impurities [15]. However, it is very likely that carbon makes up a large majority of each CNT sample. This is significant for Sample 2, as it suggests that its high mass and Faradaic efficiency are not due to contamination from the electrolysis procedure.

Thermal Analysis

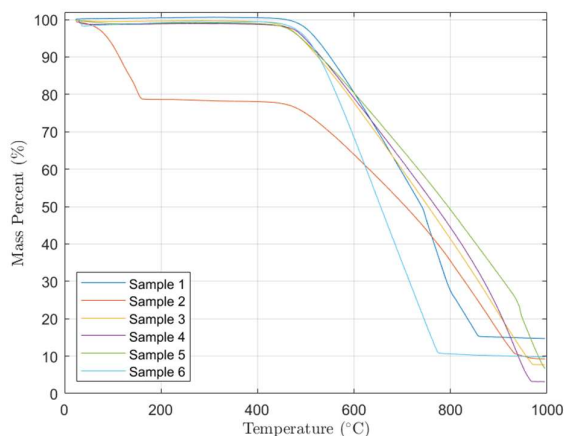


Figure 9: Thermogravimetric Analysis (TGA) results for samples from all Experiments 1-6, illustrating their thermal degradation profiles through the change in remaining mass percent as temperature is increased from 30 to 1000 °C

As shown in Figure 9, for TGA most samples demonstrate a similar trend of constant mass percent at 100% until the temperature reaches 450 °C. After this point, all samples demonstrate somewhat linear decreases of mass % as temperature increases further. Most samples then plateau out again, albeit at a range of temperatures. The plateau until 450 °C may indicate that little to no amorphous carbon material is present as literature indicates they can be expected to begin oxidising away from 350 °C [31], which would result in decreasing mass percent. The mass percentage decrease observed from ~420 °C is indicative of the presence of 'raw' MWCNTs as they will undergo oxidation from this temperature [32] to produce carbon monoxide or dioxide. There is also potential evidence for the presence of graphitised MWCNTs since oxidation occurs over a larger temperature range than otherwise expected [32].

This large range could also be due to carbon nano-onion (CNO) agglomerations being present as they have a higher oxidation onset temperature of 520 °C [33]. Potentially this wider temperature range of oxidation could also be due to the high heating rate of 10 K/min allowing mass transfer limitations to have an effect during testing [34].

Interestingly, the sample from Experiment 6 (iron & nickel) demonstrates an oxidation that completes at the lowest temperature and so occurs quickest. This may be indicative of differing CNT properties between samples, such as nanotube length, the presence of defects or metal particles, or could be more likely attributed to differences in the mass transfer limitations during the STA process, resulting from how densely packed the samples are [34]. It is likely that Sample 6 was less compacted for analysis and thus completed faster.

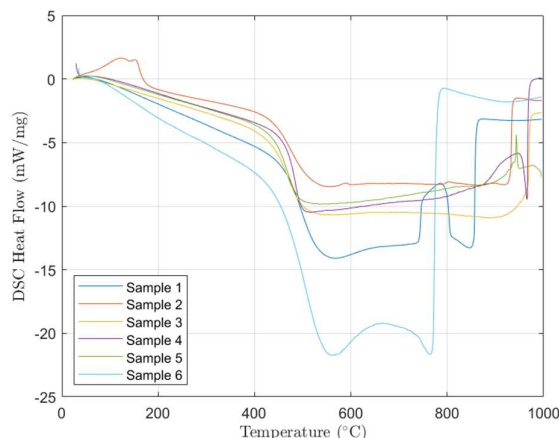


Figure 10: Differential Scanning Calorimetry (DSC) results for samples from all Experiments 1-6, illustrating how heat flow, relative to a reference in crucible, varies as temperature is increased from 30 to 1000 °C. These results were obtained simultaneously with those in Figure 9

The DSC Analysis conducted, seen in Figure 10, closely matches the observed trends in mass % over the temperature range. There is a notable decrease in the heat flow for all samples after 450 °C to a more negative value, which indicates an exothermic process is occurring. The oxidation of CNTs through combustion is exothermic and so this result strongly reinforces the previous conclusion that CNTs make up most of the carbon material within the samples. Sample 6's noticeably more negative heat flow value is likely due to the reduced mass transfer limitation to combustion as previously mentioned.

Sample 2 (iron & Inconel 625) demonstrates unusual behaviour from the start temperature of 30 °C to ~180 °C. It uniquely demonstrates a decrease in mass percent to roughly 80% of the original mass, before subsequently plateauing out and following a similar trend to the other samples. Interestingly, Sample 2's DSC analysis from 30 – 180 °C is also different from the other samples in Figure 10, with a small positive heat flow that is indicative of an endothermic process, potentially the evaporation of water or another contaminating solvent with low boiling point [35]. Given that XRD analysis does not indicate any such solvents for this sample, it is likely that any solvent contamination occurred during the analytical testing for this sample, and not during synthesis itself.

After the completion of the analysis, noticeable residues were observed left over within the testing

crucible. They had varied colours, with black, green and brick red residues observed. These residues correspond to the final plateaus in mass % at high temperatures and are likely metallic impurities, like those mentioned previously with respect to XRD, that do not melt or burn away at 1000 °C.

SEM Imaging

The SEM images produced confirm the successful synthesis of CNTs for samples from all experiments, although some differences between the obtained nanostructures can be inferred. Notably some samples such as those from Experiment 3 (Figure 13) and Experiment 4 (Figure 14) appear to be comprised of shorter, less densely packed nanotubes, with very noticeable globular structures present too. These globular structures may be agglomerations of carbon nano-onions, which are a similar carbon nanomaterial to CNTs that are also known to be synthesised during the molten salt electrolysis of carbonates [19]. Other samples such as those from Experiments 1 and 6 (see Figures 11, 12 & 15) appear to be comprised of longer, more numerous nanotubes. The length of obtained nanotubes is likely linked directly to the electrodes used during electrolysis [36], while the number and density of nanotube formation could be due to the availability of migrated nickel nucleation sites on the cathode. It then follows that samples from Experiments 3 and 4 could be comprised of fewer, shorter CNTs as the galvanised steel's zinc layer may inhibit iron and nickel availability for CNT nucleation. This also explains why samples from Experiments 1 and 6 have more plentiful CNTs formed, as the exposed iron cathode could more readily enable nucleation, and in the absence of the zinc layer, nickel could be deposited more readily to further encourage nucleation to occur.

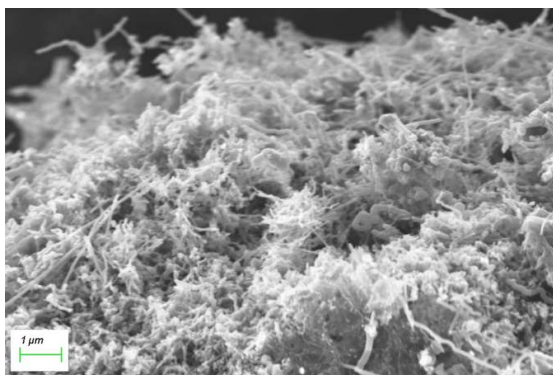


Figure 11: SEM image depicting a sample of Experiment 1's CNT product at 45.48 K $\times$ magnification

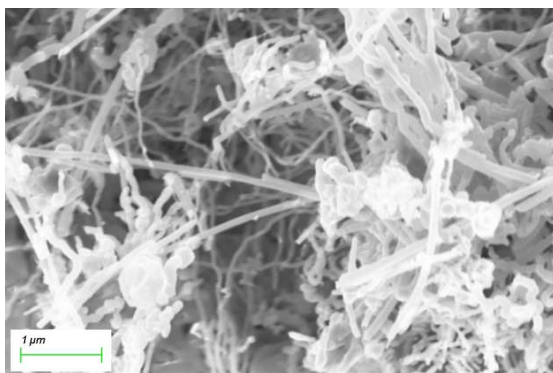


Figure 12: SEM image depicting a sample of Experiment 1's CNT product at 84.87 K $\times$ magnification

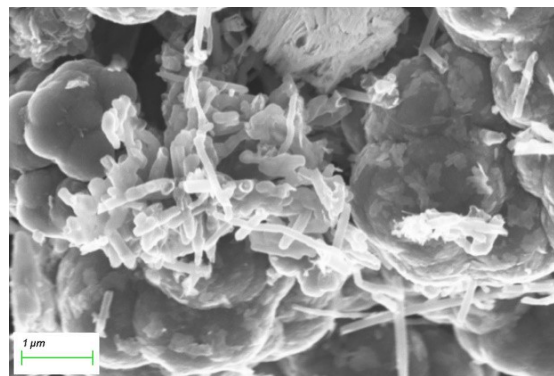


Figure 13: SEM image depicting a sample of Experiment 3's product, with short CNTs and possible CNO agglomerations at 78.20 K $\times$ magnification

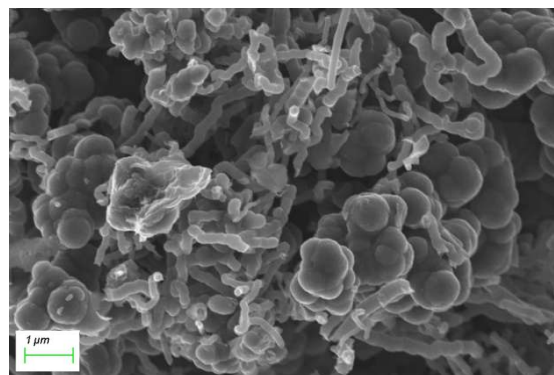


Figure 14: SEM image depicting a sample of Experiment 4's product, with short CNTs and possible CNO agglomerations at 50.34 K $\times$ magnification

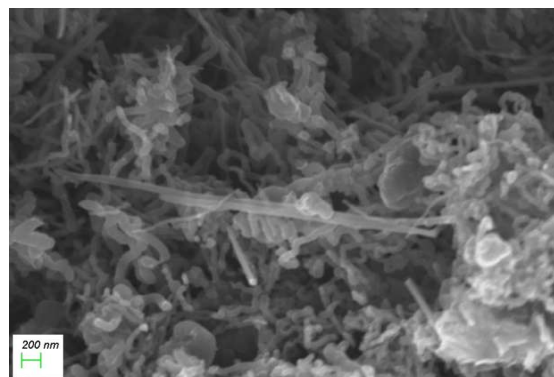


Figure 15: SEM image depicting a sample of Experiment 6's long CNT product at 101.02 K $\times$ magnification

5.0 Conclusion and Outlook

All electrode combinations utilised successfully resulted in the production of the desired highly crystalline multi-walled CNTs (MWCNTs). The generation of CNTs is assured based upon both analytical and imaging techniques, with significant amorphous carbon generation unlikely. The MWCNTs produced have ordered layers and have been proven to be as crystalline as those produced with more conventional methods, although some defects are present. Economic analysis indicates that our products would have associated electrical energy costs of 9-18 £/kg_{CNT}, significantly lower than the 120 £/kg_{CNT} energy cost that would be associated with the conventional CCVD method. Additionally, there is evidence for the production of

carbon nano-onions (CNOs) which have a similar sp^2 hybridised multi-layered structure.

With the current experimental setup, it was found that pure iron outperforms galvanised steel as the cathode. Evidence suggests that the zinc coating may interfere with the crystalline structure of CNTs and increase structural defects. The zinc's apparent ineffectiveness as a catalyst for growth likely contributes to a reduction in the yield of CNTs obtained, by limiting the availability of iron's nucleation sites. SEM imaging reaffirms these conclusions as CNTs produced using galvanised steel visually appear shorter, although CNOs are observed more frequently. Additionally, during the synthesis of paired experiments with identical anodes, it was observed that the iron cathode consistently results in sustained lower potentials and thus resistances. This also means that iron cathodes have a higher energy efficiency.

Similarly, pure nickel appears to out-perform the alloys of Inconel 625 and nichrome as an anode when high amounts of low-defect nanotubes are desired, particularly due to its possible higher energy efficiency. Our results showed that during the activation period, the potential measured is determined near solely by anodic composition, although chromium may cause fluctuations to be observed. Using nickel as an anode may involve a lower resistance reduction reaction due to the absence of other metals, such as chromium. While there was no strong observed trend between nickel content's impact on CNT defects or yield, the experimental run with pure nickel and iron was identified as the most successful for balancing the desired CNT properties.

All Faradaic efficiencies found over the course of experimentation appear to be quite low; this is likely a combined result of side-reaction and material losses due to experimental error. Experiment 2 resulted in an unusually high yield and an unusual STA result which was hard to explain when contextualised with other analyses. The STA performed could be improved through a lowered heating rate and consistently packed samples to reduce mass transfer disparities and in particular the testing of Sample 2 should be repeated to help explain its differences. Further repetitions of all of the experimental runs undertaken would also prove beneficial in ensuring the results obtained are reliably reproducible and conclusions are accurate. Additionally, XRD analysis could be improved for future experiments by using specialised software to allow for the easier interpretation of spectra and avoid the mischaracterisation of small peaks. It could also be paired with complementary EDX analysis to allow for a deeper, quantitative understanding of product composition.

To improve product Faradaic efficiencies, a weighing-by-difference methodology could be incorporated for future experimentation to reduce material loss impact, or the methodology could incorporate weighing the filter paper to quantify these material losses for error analysis. Experimentation with reduced operating temperatures could also be conducted to discourage the side reaction, provided these temperatures are maintained above the melting point of the electrolyte. Since chromium does not appear to cause increased defects, as evidenced by the lack of observed trend between disorder and anodic nickel content, stainless steel cathodes could be investigated as a suitable corrosion-resistant, iron-containing cathode.

A system allowing CO_2 to be bubbled into the electrolyte could also be introduced, as this would

regenerate the carbonate ions within the electrolyte, enabling longer synthesis periods. This would also facilitate further investigation into the feasibility of this technology as a CCU method.

Acknowledgements

We would like to express our gratitude to Prashant Acharya for his continuous guidance and support throughout the course of the project.

References

- [1] Friedlingstein, Pierre, et al. *Global Carbon Budget 2025*. Nov. 2025, <https://doi.org/10.5194/essd-2025-659>.
- [2] IPCC. "Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]. IPCC, Geneva, Switzerland." *Climate Change 2023 Synthesis Report*, July 2023, pp. 35–115, <https://doi.org/10.59327/ipcc/ar6-9789291691647>.
- [3] Liu, Nan, et al. "Unlocking Carbon Capture and Storage Potential: Policy Incentives, Economic Challenges, and Infrastructure Integration for CO_2 Transport." *CHAIN*, vol. 2, no. 3, Institute of Electrical and Electronics Engineers (IEEE), Sept. 2025, pp. 211–26, <https://doi.org/10.23919/chain.2025.000015>.
- [4] McKinsey & Company. *Decarbonization of industrial sectors: The next frontier*. June 2018.
- [5] De Volder, Michael F. L., et al. "Carbon Nanotubes: Present and Future Commercial Applications." *Science*, vol. 339, no. 6119, Jan. 2013, pp. 535–39, <https://doi.org/10.1126/science.1222453>.
- [6] Dr Conor O'Brien. "Billion Dollar Carbon Nanotube Market Continues to Grow." *IDTechEx*, 17 June 2025, www.idtechex.com/en/research-article/billion-dollar-carbon-nanotube-market-continues-to-grow/33366. Accessed 17 Nov. 2025.
- [7] Deng, Sijie, et al. "Effect of Industrial Multi-Walled Carbon Nanotubes on the Mechanical Properties and Microstructure of Ultra-High Performance Concrete." *Cement and Concrete Composites*, vol. 156, Feb. 2025, <https://doi.org/10.1016/j.cemconcomp.2024.105850>.
- [8] Sivakumar, Mani, and Sakhon Ratchahat. "Recent Progress in Carbon Nanotubes Production via Catalytic Chemical Vapour Deposition." *Journal of Analytical and Applied Pyrolysis*, vol. 189, Aug. 2025, p. 107093, <https://doi.org/10.1016/j.jaap.2025.107093>.
- [9] Dipta, I. Ketut Rai Asmara, and Chan Woo Lee. "Recent Advances and Perspectives in Carbon Nanotube Production from the Electrochemical Conversion of Carbon Dioxide." *Journal of CO_2 Utilization*, vol. 82, Apr. 2024, p. 102745, <https://doi.org/10.1016/j.jcou.2024.102745>.
- [10] Dipta, I. Ketut Rai Asmara, and Chan Woo Lee. "Electrode Materials for Electrochemical Synthesis of Carbon Nanotube from CO_2 : Surface Properties, Mechanisms, and Perspectives." *Applied Surface Science Advances*, vol. 29, Elsevier, Aug. 2025, p. 100823, <https://doi.org/10.1016/j.apsadv.2025.100823>.
- [11] Liu, Xinye, et al. "Controlled Growth of Unusual Nanocarbon Allotropes by Molten Electrolysis of CO_2 ." *Catalysts*, vol. 12, no. 2, Jan. 2022, p. 125, <https://doi.org/10.3390/catal12020125>.
- [12] Chery, D., et al. " CO_2 Electrochemical Reduction into CO or c in Molten Carbonates: A Thermodynamic Point of View." *Electrochimica Acta*, vol. 160, Apr. 2015, pp. 74–81, <https://doi.org/10.1016/j.electacta.2015.01.216>.

- [13] Li, Zhida, et al. "Carbon Nanotubes Synthesis from CO₂ Based on the Molten Salts Electrochemistry: Effect of Alkaline Earth Carbonate Additives on the Diameter of the Carbon Nanotubes." *Journal of the Electrochemical Society*, vol. 166, no. 10, 2019, pp. D415–20, <https://doi.org/10.1149/2.0861910jes>.
- [14] Ijije, Happiness V., et al. "Electro-Deposition and Re-Oxidation of Carbon in Carbonate-Containing Molten Salts." *Faraday Discussions*, vol. 172, Royal Society of Chemistry, Jan. 2014, pp. 105–16, <https://doi.org/10.1039/c4fd00046c>.
- [15] Laasonen, Emma, et al. "The Effect of Metal Dissolution on Carbon Production by High-Temperature Molten Salt Electrolysis." *Journal of CO₂ Utilization*, vol. 69, Elsevier BV, Mar. 2023, pp. 102390, <https://doi.org/10.1016/j.jcou.2023.102390>.
- [16] Dey, Gangotri, et al. "How Does an Amalgamated Ni Cathode Affect Carbon Nanotube Growth? A Density Functional Theory Study." *RSC Advances*, vol. 6, no. 32, Royal Society of Chemistry (RSC), 2016, pp. 27191–96, <https://doi.org/10.1039/c6ra03460h>.
- [17] PubChem. "Zinc." *Nih.gov*, PubChem, 2024, pubchem.ncbi.nlm.nih.gov/compound/Zinc.
- [18] Kavunga, Sunil, et al. "In Situ Characterization of Galvanized Low-Alloyed Steels with High-Temperature Cyclic Voltammetry during Annealing." *Electrochimica Acta*, vol. 424, Elsevier BV, Aug. 2022, p. 140653, <https://doi.org/10.1016/j.electacta.2022.140653>.
- [19] Liu, Xinye, et al. "Carbon Nano-Onions Made Directly from CO₂ by Molten Electrolysis for Greenhouse Gas Mitigation." *Advanced Sustainable Systems*, vol. 3, no. 10, Wiley, July 2019, <https://doi.org/10.1002/adsu.201900056>.
- [20] Ren, Jiawen, et al. "One-Pot Synthesis of Carbon Nanofibers from CO₂." *Nano Letters*, vol. 15, no. 9, Aug. 2015, pp. 6142–48, <https://doi.org/10.1021/acs.nanolett.5b02427>.
- [21] ---. "Transformation of the Greenhouse Gas CO₂ by Molten Electrolysis into a Wide Controlled Selection of Carbon Nanotubes." *Journal of CO₂ Utilization*, vol. 18, Mar. 2017, pp. 335–44, <https://doi.org/10.1016/j.jcou.2017.02.005>.
- [22] "Meter Loan Pool." *Imperial College London*, 2024, www.imperial.ac.uk/sustainability/our-campuses/meter-loan-pool/. Accessed 8 Dec. 2025.
- [23] Eilers, Paul H. C., and Hans Boelens. *Baseline Correction with Asymmetric Least Squares Smoothing*. Oct. 2005.
- [24] Pal, Sunil K., et al. "Importance of Cr₂O₃ Layer for Growth of Carbon Nanotubes on Superalloys." *Carbon*, vol. 48, no. 3, Mar. 2010, pp. 844–53, <https://doi.org/10.1016/j.carbon.2009.10.038>.
- [25] Fugetsu, Bunshi, et al. "Electrical Conductivity and Electromagnetic Interference Shielding Efficiency of Carbon Nanotube/Cellulose Composite Paper." *Carbon*, vol. 46, no. 9, Aug. 2008, pp. 1256–58, <https://doi.org/10.1016/j.carbon.2008.04.024>.
- [26] Pimenta, M. A., et al. "Studying Disorder in Graphite-Based Systems by Raman Spectroscopy." *Phys. Chem. Chem. Phys.*, vol. 9, no. 11, 2007, pp. 1276–90, <https://doi.org/10.1039/b613962k>.
- [27] Saito, R., et al. "Raman Spectroscopy of Graphene and Carbon Nanotubes." *Advances in Physics*, vol. 60, no. 3, June 2011, pp. 413–550, <https://doi.org/10.1080/00018732.2011.582251>.
- [28] Delhaes, P., et al. "A Comparison between Raman Spectroscopy and Surface Characterizations of Multiwall Carbon Nanotubes." *Carbon*, vol. 44, no. 14, Nov. 2006, pp. 3005–13, <https://doi.org/10.1016/j.carbon.2006.05.021>.
- [29] Garzon-Roman, A., et al. "Evaluation of Synthesis Time in the Growth of Vertical-Aligned MWCNTs by Spray Pyrolysis." *Materials Characterization*, vol. 203, Elsevier BV, Sept. 2023, p. 113105, <https://doi.org/10.1016/j.matchar.2023.113105>.
- [30] Reza Teimuri-Mofrad, et al. "Synthesis, Characterization and Electrochemical Study of Carbon Nanotube/Chitosan-Ferrocene Nanocomposite Electrode as Supercapacitor Material." *Journal of Electronic Materials*, vol. 48, no. 7, Springer Science+Business Media, May 2019, pp. 4573–81, <https://doi.org/10.1007/s11664-019-07226-2>.
- [31] Shi, Zujin, et al. "Purification of Single-Wall Carbon Nanotubes." *Solid State Communications*, vol. 112, no. 1, Pergamon, Aug. 1999, pp. 35–37, [https://doi.org/10.1016/S0038-1098\(99\)00278-1](https://doi.org/10.1016/S0038-1098(99)00278-1).
- [32] Mahajan, Amit, et al. "Studies on the Thermal Decomposition of Multiwall Carbon Nanotubes under Different Atmospheres." *Materials Letters*, vol. 90, Jan. 2013, pp. 165–68, <https://doi.org/10.1016/j.matlet.2012.08.120>.
- [33] Plonska-Brzezinska, Marta E., et al. "Post-Modification by Low-Temperature Annealing of Carbon Nano-Onions in the Presence of Carbohydrates." *Carbon*, vol. 67, Elsevier BV, Feb. 2014, pp. 304–17, <https://doi.org/10.1016/j.carbon.2013.09.093>.
- [34] Zapata, John A., et al. "Influence of Tapped Density on the Degradation Profile of Multiwall Carbon Nanotubes." *Thermochimica Acta*, vol. 654, Elsevier BV, May 2017, pp. 140–45, <https://doi.org/10.1016/j.tca.2017.05.014>.
- [35] Trairat Muangthong-on, et al. "TG-DSC Study to Measure Heat of Desorption of Water during the Thermal Drying of Coal and to Examine the Role of Adsorption of Water Vapor for Examining Spontaneous Heating of Coal over 100 °C." *Energy & Fuels*, vol. 31, no. 10, American Chemical Society, Sept. 2017, pp. 10691–98, <https://doi.org/10.1021/acs.energyfuels.7b01836>.
- [36] Liu, Xinye, et al. "Controlled Transition Metal Nucleated Growth of Carbon Nanotubes by Molten Electrolysis of CO₂." *Catalysts*, vol. 12, no. 2, Jan. 2022, p. 137, <https://doi.org/10.3390/catal12020137>.

A Systematic Techno-economic Review of Carbon Capture and Storage Technologies in the Global Cement Industry

Alya Dania Adiraputri, Jing Qiao Soo

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Policy mechanisms such as the EU and UK's ETS serve as direct economic incentives for energy-intensive industries such as cement to reduce carbon emissions. This study provides an updated systematic review of techno-economic studies of carbon capture and storage applied to clinker and cement production. In particular, different design configurations within oxyfuel combustion, calcium looping (CaL), amine scrubbing, and membrane separation are discussed. The comparative analysis is conducted through metrics such as Cost of Clinker, Cost of CO₂ Avoided, and Specific Primary Energy Consumption for CO₂ Avoided to understand the energy penalties associated with each technology type, and the contribution of capital and operating expenses with its associated assumptions. The narrowing gap between ETS prices and the Cost of CO₂ Avoided demonstrates that CCS projects are approaching economic viability. There is a disparity observed within academic literature and industry, with 65% of identified commercial projects pursuing oxyfuel or amine scrubbing technologies, despite calcium looping being identified as having the lowest CAC and a low SPECCA.

Key Words: Systematic review, cement industry, clinker, calcium looping, amine scrubbing

1. Introduction

In 2024, the total net anthropogenic greenhouse gas emissions reached 57.7 GtCO₂, a 2.3% increase from 2023. The increase in annual emissions rate was quadruple the rate in the 2010s and is similar to the 2000s (United Nations, 2025). The global cement industry is responsible for 7% of total CO₂ emissions and is not on track for net zero by 2050 (European Commission, 2023). While there has been a decrease in total emissions in recent years due to a reduction in global demand, direct CO₂ emissions intensity remains unchanged (IEA, 2025). Hence, more efforts are needed to decarbonise the cement industry.

Cement production begins from extracting raw materials (limestone, clay and shale) and crushing them into smaller pieces. These are then transported to the cement plant, where they are ground to fine powders. The core and most energy-intensive unit of cement production is the rotary kiln, where fine raw materials are heated to around 1450°C to chemically transform into clinker. Equation 1 describes the decarbonation of limestone (CaCO₃), which forms the main constituent of clinker. After leaving the kiln, the clinker is cooled and ground into fine powder with the addition of Gypsum, to produce the final cement product. To reach the required high temperatures in the kiln for clinker formation, fossil fuels are combusted, which in turn releases CO₂ into the atmosphere (Barbhuiya et al., 2024).



One promising method to decarbonise cement production is to incorporate carbon capture and storage (CCS) technologies into the process line. However, production costs for early commercial cement plants using CCS are estimated to be 75-150% higher than today's conventional plants (IEA, 2025). With rising carbon taxes and new legislation across

different regions, it is predicted that it may reach higher levels than the cost per tonne of carbon emitted in cement production. The European Union's Carbon Border Adjustment Mechanism (CBAM) builds upon the EU Emissions Trading System (ETS) to prevent carbon leakage by phasing in carbon prices for importers and phasing out free allocation for domestic producers (Bloomberg, 2025). The UK has also implemented an ETS policy to incentivise investments in decarbonization, although it is at a discount to the EU's at 55 £/tCO₂ (Chapman, 2025). These policies serve as additional motivation to invest in decarbonising the cement industry, particularly within Europe.

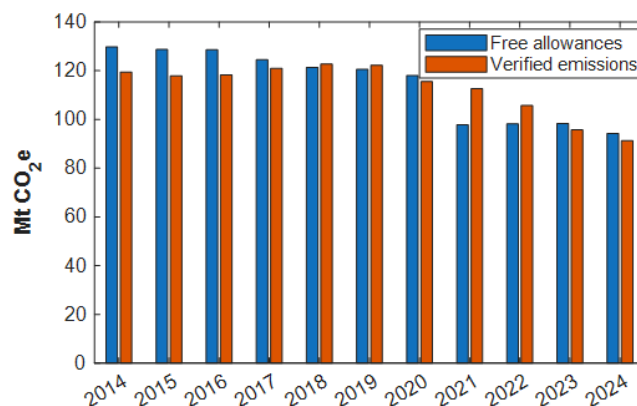


Figure 1: Historical allocations and emissions in the cement clinker sector under EU ETS.

Figure 1 demonstrates that the cement industry has been able to reduce emissions within the allocated allowance in recent years. This has been achieved through a combination of reducing clinker production, utilising supplementary cementitious materials and alternative fuels (CEMBUREAU, 2024). However, further decarbonisation is required

to achieve net-zero production, which can only be achieved with CCS solutions.

2. Background

2.1 Overview of CCS technologies discussed

This review focuses on 4 primary technologies for implementing carbon capture within the clinker production process, namely oxyfuel combustion, calcium looping, amine scrubbing, and membrane separation.

Oxyfuel combustion utilises an Air Separation Unit (ASU) to provide an O_2 stream with recycled flue gas into the precalciner and rotary kiln. The exhaust is therefore composed of CO_2 rich stream and H_2O , requiring only water and impurity removal in the Compression and Purification Unit (CPU) before transport and storage (Rolfe et al., 2018).

Calcium looping exploits the reversible reaction above in Figure 1, composed of two interlinked circulating fluidised bed reactors in the form of the carbonator and calciner. The exothermic backwards/reverse reaction occurs in the carbonator, where CO_2 rich flue gases from the precalciner and rotary kiln reacts with the calcium-based sorbent to form $CaCO_3$. The CO_2 lean flue gases then leaves the system, and the $CaCO_3$ enters the calciner, where the endothermic forward reaction occurs. Heat is provided by combustion in an oxygen rich environment, also provided by the ASU, and a CO_2 stream leaves to the CPU, with regenerated sorbent cycled back to the carbonator, or purged as feed to the rotary kiln (Atsonios et al., 2015).

Amine scrubbing utilises a two column configuration with typically a MEA (or other amine-based) solvent. The CO_2 rich flue gases enter the absorber column, where it contacts the solvent in counter-current flow. As the CO_2 is adsorbed to the solvent, it flows down and exits the bottom of the column, while CO_2 lean gases are washed with water as they flow up, reducing solvent losses. The CO_2 rich solvent enters the stripping column, where heat is supplied by the condenser so a CO_2 stream is produced. The regenerated solvent exits from the reboiler bottoms, and re-enters the absorber with fresh solvent (Strojny et al., 2025).

Membrane-based separations utilise the semi-selective properties to allow one or more gases to pass through, producing a retentate (rejected fluid) and a permeate that is CO_2 rich. A combination of blowers, vacuums, and compressors creates the pressure difference used to drive the separation. Multiple membrane stage systems can be configured to optimise for CO_2 capture and purity (Scholes et al., 2014).

2.2 Previous works

Previous reviews have aimed to demonstrate gaps and weaknesses in industrial CCS research, with Leeson et al. (2017) sourcing 15 sources related to the cement industry, including both academic and grey literature. The authors identified oxy-combustion with calcium looping as the more economically favourable technology over the well-established amine-based separation, based on CAC.

Garcia et al. (2018)'s techno-economic review also aimed to homogenize studies related to CO_2 capture in the cement

industry by standardizing key technical parameters such as plant size and capacity factors, as well as applying the same economic assumptions on cost calculations. 32 cases were reviewed for the cement industry, and again, the authors observe that technologies such as calcium looping and indirect calcination offer the lowest CAC, but is highly dependent on assumptions of heat integration, and export of surplus energy to the grid.

Santos and Hanak (2022)'s review assessed both CAC and the thermodynamic performance of different technologies by utilising equivalent energy consumption ($MJ_{th}/kgCO_2$ captured). The authors' discussion focused on solid looping cycles, and in particular, calcium looping. Oxy-fuel combustion yielded the lowest equivalent energy consumption of $1.5MJ_{th}/kgCO_2$ compared to CaL's $2.2MJ_{th}/kgCO_2$, although this also comes in hand with lower capture rates.

This paper provides an updated review of the CCS technologies within the cement industry, examining the different cost drivers and temporal trends observed within academic literature. The energy penalty of each capture system is also examined through the Specific Primary Energy Consumption for CO_2 Avoided (SPECCA). A link to industry application is also drawn by making connections to carbon tax, and real world industry projects.

3. Methodology

3.1 Systematic review and screening

Search terms were constructed and combined with Boolean operators, then refined as indicated in the supplementary data. This is to find the most relevant literature for this paper across Engineering Village and EBSCO databases, resulting in 187 unique papers to be screened. An initial screening of these papers was conducted by reading their abstracts, returning 138 relevant papers. A 'relevant' article's abstract is defined as: (1) Appears to contain information on the technology, economics, or legislation of carbon capture that is highly applicable or comparable to the cement industry. (2) Appears to contain indirect information relating to technology, economics or legislation of carbon capture that is highly applicable or comparable to the cement industry. (3) Appears to contain supporting information that might reasonably be used to provide contextualising information on carbon capture implementation within the cement industry.

Literature satisfying these criteria was then holistically reviewed using a standard questionnaire detailing technology, cost, and policy surrounding CCS in the cement industry. A short technical, economic or general summary was written as part of the reviewing process, alongside the level of priority of the paper, which can be found in the supplementary information. This methodology is adapted from Leeson et al. (2017).

3.2 Data standardisation and cost escalation

The main metrics used for comparison in this study are the Cost of Clinker (Eq. 2), the Cost of CO_2 Avoided (Eq. 3), and the Specific Primary Energy Consumption for CO_2 Avoided (Eq. 3). The subscripts *reference* and *capture* refers to the

cement plant with and without a carbon capture system, and e and q refers to the CO₂ equivalent emissions and primary energy consumption, respectively. This includes direct fuel consumption and indirect energy (such as electricity).

$$COC = \frac{Ann.CAPEX + Ann.OPEX}{Ann.ClinkerProd.} \quad (2)$$

$$CAC = \frac{COC_{capture} - COC_{reference}}{e_{reference} - e_{capture}} \quad (3)$$

$$SPECCA = \frac{q_{capture} - q_{reference}}{e_{reference} - e_{capture}} \quad (4)$$

The Cost of CO₂ Avoided is used as the comparison metric rather than the Cost of CO₂ Captured, as it accounts for additional energy penalties and indirect emissions from the capture system. This metric also allows a determination of a break-even point for carbon tax or carbon market prices. However, since this cost figure is still highly dependent on economic assumptions, the SPECCA metric is more suited to quantifying energy penalties and indirect emissions.

$$Cost_{2024} = Cost_{BaseYear} \times \frac{CEPCI_{2024}}{CEPCI_{BaseYear}} \quad (5)$$

Cost escalation was also conducted to compare historical values using the same currency (USD). Currency conversion data was obtained from the International Monetary Fund (IMF, 2025), and subsequently escalated to 2024 values using data from the CEPCI (2014) via Equation 5. For a more effective comparison, costs should also be multiplied by a location factor.

3.3 Results and Demographic of Systematic Review

The demographics of the 138 papers reviewed are as follows: 32 high priority papers, and 29 medium priority papers. A total of 25 of 138 (21.7%) original literature paper consists of the techno-economics in cement focused carbon capture, 7 cost papers were excluded as they were outdated. 13 articles (9 from the 25 cost papers) comprised of information on energy demands and/or metrics. 17 papers cover the policies surrounding general and cement industry capture systems. An additional 10 public company reports were selected to understand the recent development of carbon capture projects in the cement industry. Despite search terms including "storage", there is a lack of cost or energy data to make a viable comparison. Only 3 out of 25 papers included transport and storage (T&S) considerations in cost data.

The Cost of CO₂ avoided (CAC) was the most frequently observed value across 25 papers, leading to a total of 65 data points. The distribution of data is spread across different types of CCS technologies: 43% of CAC data were for CaL, 25% for amine scrubbing, 12% for membrane-based capture, 6% for oxyfuel combustion, and the remaining 14% corresponds to other/combined technologies. A similar distribution of other available techno-economics and energy metrics can be seen for each technology type. Despite the emphasis on CaL in academia, the cement industry is currently deploying more non-CaL based CCS.

4. Results and discussion

4.1 Summary of literature cost

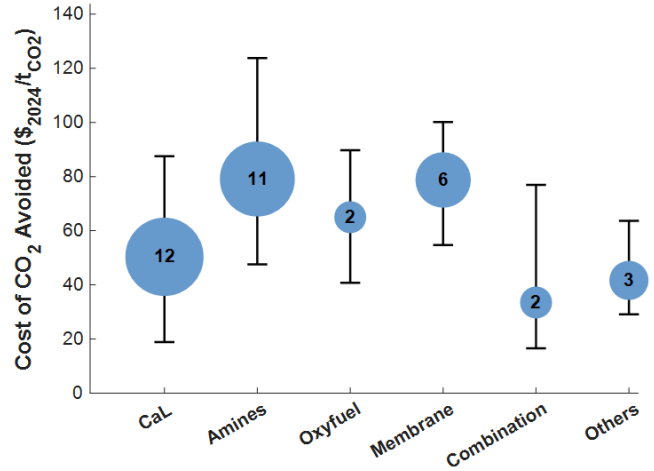


Figure 2: Mean costs of CO₂ avoided with error bars representing the range of costs found in the literature for each technology, excluding T&S. The diameter of the bubble is proportional to the number of relevant cost papers.

Avoided cost results including transport and storage (T&S) has been omitted due to data scarcity. The assumptions made varies completely, such as transportation distance (location) and recompression power in pipelines.

The relevant cost data for each capture technology in the cement industry are summarised in Figure 2. Out of the four main technologies (CaL, Amines, Oxyfuel and Membranes), CaL shows the lowest mean CAC (50.4 \$2024/tCO₂), followed by oxyfuel (65.0 \$2024/tCO₂), then membranes (78.8 \$2024/tCO₂), and lastly amines (79.1 \$2024/tCO₂). The consensus when breaking down the CAC of each process lies in the energy demands that are fulfilled using either fuel, electricity or steam.

The discussion of statistical measures is only valid for CaL and amines due to the sufficient volume of papers, resulting in 28 and 16 cost data, respectively. The disproportionate research focus in CaL indicates a research bias in exploring viable cement CCS technologies. CaL displays a lower standard deviation (16.8) compared to amines (21.2). While amine scrubbing has a higher technological readiness level (TRL) than CaL, the wider spread dictates a greater sensitivity to changes in steam production methods to satisfy regeneration duty. Cost estimates vary depending on the auxiliary boilers (fuel type) and heat waste recovery system.

The high CAC of amine is because of the high energy requirement to regenerate the solvent. Primary energy drives in amine have a need for auxiliary boilers, because unlike power plants, cement kilns do not produce sufficient heat to cover the energy requirements of the amine stripper (Garcia et al., 2018). The regeneration energy in amine scrubbing consists of the heat of reaction, sensible heat, and latent heat of vaporisation. Research on utilising other solvents than traditional MEA, such as PZ-AMP has proven to have better kinetic performance. PZ-AMP reduces regeneration energy by having a lower reaction enthalpy and higher rich load-

ing. Therefore, lowering the temperature profile across the stripper column significantly lowers the latent heat. It had a regeneration duty 39% less than MEA based capture, with a 16% reduction in CAC of 73.8 $\$/_{2024}/t_{CO_2}$ (Jiang et al., 2024). Furthermore, using a lean vapour compressor (LVC) with an optimised heat exchanger reduces the CAC by 20% yet increases plant cost by 7% compared to the standard MEA process (Aromada et al., 2022).

Membrane capture methods have the second highest mean CAC because of the substantial compression required for a high-pressure difference to operate the membrane. It is often combined with cryogenic liquefaction to remove non-condensable gases, making the whole process energy intensive (Cloete et al., 2020).

In an oxyfuel capture system, the ASU and CPU are both capital and electrical-intensive. Cost arises from high energy penalty, and retrofitting is difficult as the kiln needs to be remade to be compatible (Gardarsdottir et al., 2019).

In CaL, selling electrical surplus drives the CAC down. The range in CaL is driven by the configuration and heating types: tail-end, integrated, direct or indirect heating. Tail-end and indirect schemes are more suitable for brownfield plants, whereas the other two are better for greenfield plants. The cost of the CO_2 capture unit for the tail-end system is 11M€ more expensive than the integrated system (De Lena, Spinelli, et al., 2019). Assuming the cement plant cost of the integrated system is fully amortised, the upfront investment cost for integrated system is lower as it considers only the cost of CO_2 capture plant. On the other hand, tail-end plant considers both the cement and CO_2 capture plant, leading to a higher upfront cost.

Besides the four main CCS technologies, cost for the combination and other types of technology was found. However, it is not viable to compare the means due to the varying technologies within each type. Hence, it is more logical to understand the cost individually. Under combined technologies, Strojny et al. (2025) reported a combination of amine scrubbing and CaL technology, leading to a maximum cost at 77.0 $\$/_{2024}/t_{CO_2}$. It is 30% cheaper than the CAC of MEA capture reported from the same paper, but with a 3% increase in COC. Wu et al. (2023) discusses 3 methods of utilising CO_2 capture via CaL on-site to produce ammonia, urea, and methanol. The difference in the 3 methods is the dependence on the expensive green electricity grid and technique to produce the valorised product. Relying wholly on the grid and a polymer electrolysis membrane leads to the lowest cost reported of 16.6 $\$/_{2024}/t_{CO_2}$ due to the sale of products. This competes with the prices of conventional CaL, highlighting a viable hybrid approach. The author acknowledges that integrating a capture unit is superior when carbon tax is above $\$/_{55}/ton$.

In other technologies, 3 cost papers highlight 3 different approaches. The cost of 63.6 $\$/_{2024}/t_{CO_2}$ is from a chilled ammonia process using the same method as amine scrubbing (Gardarsdottir et al., 2019). From the same literature, CAC is 12% cheaper than MEA-based capture due to lower steam re-

generation requirements. Though there is a trade-off with the CAPEX being roughly double that of MEA capture. Cloete et al. (2020) utilised a swing adsorption reactor that relies completely on electricity, with the same CAC as the mean cost of CaL. The author argues that it is easier to retrofit than CaL.

Nelson et al. (2017) conducted a pilot-scale experiment using a solid sorbent in a thermal swing absorption-desorption process, as part of HeidelbergCement's test study for CCS in 2013. The study concluded in 2017, with a reported range of 29.1 – 34.5 $\$/_{2024}/t_{CO_2}$. Although a promising economic result, the pilot was regularly unstable during the test time, hence further development in process stability is required to become technically viable (Gassnova, 2017).

These unconventional technologies categorised under 'other' report a lower CAC than mature amine-based capture systems, demonstrating potential economic viability and warranting further research.

4.2 Summary of SPECCA from literature

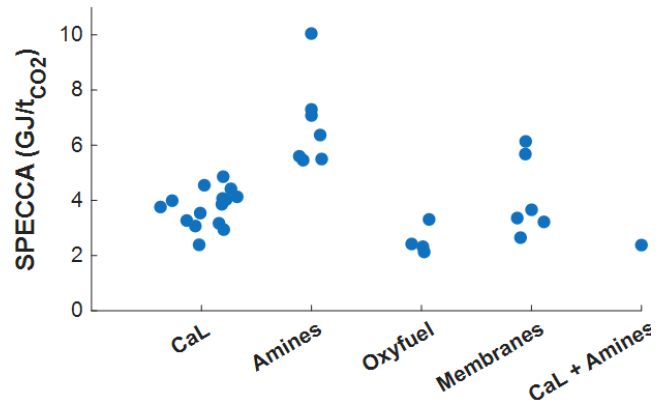


Figure 3: Swarm chart of SPECCA across different technology types.

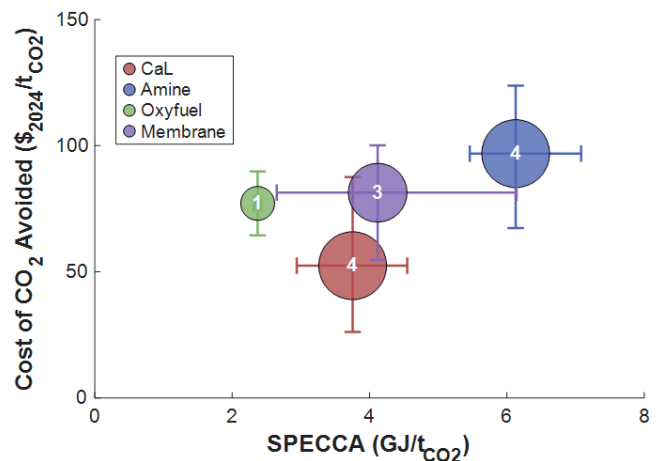


Figure 4: Mean CAC against SPECCA on different capture technology. Vertical and horizontal error bars represent the range of CAC and SPECCA respectively, with the diameter of the bubble reflecting number of literature paper.

The trend in SPECCA shown in Figure 3 generally follows the same trend of CAC in Figure 2. Rolfe et al. (2018) reports that the fuel consumption in CaL is significantly higher

than the electrical demands in oxyfuel, yet CaL (2.39 VS 3.31 GJ/tCO₂) has a lower SPECCA. When there is sufficient electrical surplus, all cost sources discussing CaL assume selling of electricity to the grid. Exporting electricity from CaL offsets the net energy requirement and is what makes the technology attractive.

Unlike other solvent/sorbent-based technologies, oxyfuel combustion has no regeneration step, avoiding the high thermal regeneration penalties (Jiang et al., 2024). Instead, the energy penalty is primarily electrical and is driven by the ASU. It has the lowest SPECCA with a range of 2.32 – 2.42 GJ/tCO₂, with corresponding CAC of 61.3 – 85.4 \$₂₀₂₄/tCO₂. Total oxyfuel combustion corresponds to the lower range, whereby oxyfuel is used to fire both the pre-calciner and rotary kiln, achieving 93% decarbonation efficiency. The higher range is from a partial oxyfuel combustion process, where it only fires the pre-calciner, achieving a 62% decarbonation efficiency (Cormos, 2022). This demonstrates a significant amount of CO₂ is emitted from heating the kiln with fossil fuels.

Despite oxyfuel exhibiting a generally lower SPECCA, it has a higher CAC than CaL, as seen in Figure 4. A hybrid approach in CaL was investigated using waste heat recovery to produce hydrogen rather than using oxyfuels to fire the kiln. This led to a reduction in SPECCA and CAC by 27% and 20 % respectively, again highlighting the high cost of ASU (Zheng et al., 2025).

Amine-based capture has the greatest energy penalty as reflected in the highest mean of SPECCA of 6.13 GJ/tCO₂. Primary energy drivers in amine scrubbing have a need for auxiliary boilers, as cement kilns do not produce sufficient heat to cover the energy requirements of the stripper column. The regeneration energy in amine scrubbing consists of the heat of reaction, sensible heat, and latent heat of vaporisation. PZ-AMP reduces regeneration energy by having a lower reaction enthalpy and higher rich loading. Therefore, it lowers the temperature profile across the stripper column, significantly lowering latent heat (Jiang et al., 2024).

Membrane technology exhibits a large range in both variables, with a SPECCA of 2.65 – 6.14 GJ/tCO₂ and CAC of 54.7 – 100.1 \$₂₀₂₄/tCO₂. The discrepancy is due to each data point exploring different types of membrane technology. Generally, a higher number of membrane stages leads to a lower SPECCA and CAC (up to an optimal point), as it minimises the electrical compression work needed. For a triple stage polymeric membrane/ Cormos (2022) reported the lowest SPECCA of 2.38 GJ/tCO₂, whereas a two stage metal organic framework membrane yields a SPECCA range of 3.36 -3.66 GJ/tCO₂ (Kvamsdal et al., 2021). A single-stage facilitated transport membrane consuming both electricity and steam corresponds to the highest SPECCA of 6.14 GJ/tCO₂ (Ferrari et al., 2021).

4.3 Sensitivity of fuel and electricity origin and price

Assumptions behind the purchase and source of fuel and electricity contribute to indirect emissions, hence affecting SPECCA and costs. With membranes and oxyfuel technol-

ogy being power-intensive, their CAC is proportional to price and carbon intensity of the electrical supply (Lee and Kim, 2023), such as detailed in Table 1 below.

Table 1: Selection of grid electricity emission factors and locations observed from literature reviewed, in kgCO₂/MWh.

EU	US	China	Uzbekistan
238-262	390-640	856	617

Tail-end CaL configuration consumes more fuel due to a double calcination process, but it can export more electricity, compared to integrated CaL systems (De Lena, Spinelli, et al., 2019). The sold electricity credit is what drives the CAC down. Sensitivity analysis conducted by De Lena, Spinelli, et al. (2019), shows that this configuration is the cost-effective option when electricity prices are high, but also remains highly sensitive to fuel price fluctuations. Conversely, integrated systems are more susceptible to changes in electricity prices. Although integrated configuration is thermodynamically more favourable as it achieves a lower SPECCA range (3.17 - 3.27 GJ/tCO₂) compared to tail-end systems (3.76 - 4.42 GJ/tCO₂), they can incur a higher CAC.

The variance in these SPECCA values stems from using two different energy sources: a pulverised coal power plant and the EU-28 energy mix. The latter has 3 times less carbon emission than coal (De Lena, Spinelli, et al., 2019). While using the EU-28 mix reduces the SPECCA for integrated systems (3.17 GJ/tCO₂), it contrarily increases the SPECCA for tail-end CaL (4.42 GJ/tCO₂). The contrasting relationship found for tail-end is because selling electricity to a "cleaner" grid reduces the credits gained, reducing the net CO₂ avoided.

For amine-based capture systems, the contribution of fuel consumption in the auxiliary boiler dominates at about 90% of the SPECCA (Ferrario et al., 2024). Therefore, increasing fuel prices will correspondingly increase CAC (Gardarsdottir et al., 2019). The type of fuel contributes to the avoided cost as well. Five cost data collected as a subset of Figure 2 using strictly natural gas as fuel sits below the mean cost (47.6 – 76.7 \$₂₀₂₄/tCO₂), whereas using coal sits above the mean (48.4 – 123.8 \$₂₀₂₄/tCO₂). Additionally, coal-fired systems result in a SPECCA of 7.3 GJ/tCO₂, whereas natural gas showed 5.5 GJ/tCO₂ (Ferrario et al., 2024). This is due to a lower CO₂ emission and impurities in natural gas (Cormos C. and Cormos A., 2015). Strojny et al. (2025) discusses utilising biomass, resulting in a 50% increase in avoided cost over the reference MEA plant using natural gas. Although this is economically uncompetitive, it represents an opportunity for BECCS if CO₂ from the biomass boiler can be captured and stored, resulting in net negative emissions.

4.4 Cost breakdown

Due to the wide range of design decisions and assumptions taken across the reviewed literature, a further subset of papers following the CEMCAP project methodology was examined. This allows for greater standardisation and understanding of cost drivers for each of the CCS technologies. As there is greater availability in data, the breakdowns of cost drivers

are done in terms of Cost of Clinker, rather than CAC. In particular, the Cost of Clinker is broken down into CAPEX, fixed OPEX, and variable OPEX contributions.

Fixed OPEX is a contribution of maintenance, insurance/tax, and labour costs, typically assumed as a fixed percentage of total plant costs. Variable OPEX includes all other operating expenses such as fuel, electricity, and steam costs.

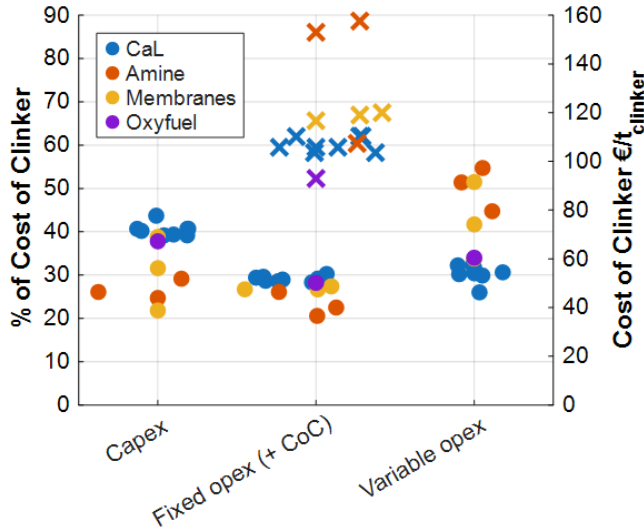


Figure 5: Cost of clinker (denoted by 'x') reported by literature following the CEMCAP project methodology, with corresponding % shares of cost (denoted by 'o') by primary technology types. Data taken from 5 unique literature.

Of the four technologies, calcium looping and oxyfuel can be seen to exhibit similar cost contributions in Figure 5. Both technologies have the greatest % share of CAPEX, with the ASU and CPU being the most significant cost components. In some cases, the ASU and CPU represent more than 50% of CAPEX for CaL configurations (De Lena et al, 2019).

The use of an indirect heating CaL configuration has been proposed to replace the need for an ASU and CPU, by utilising heat pipes to transfer thermal energy to the calciner from a separate air-fired combustor. A study by Yin et al., 2024 found that SPECCAs (in MJ/tclinker) obtained for the IH-CaL configurations were lower than those of traditional CaL by 49-60%, representing a significant opportunity for energy and cost savings.

Amine and membrane-based technologies represent the highest Cost of Clinker, and with the highest cost contributions from variable OPEX. This is driven by steam and energy costs from the reaction heat and latent heat requirements in amine-based systems, as discussed previously. In contrast, standard membrane systems are primarily driven by electricity costs to power compressors, blowers, and vacuums to create the pressure difference used to drive separation. Steam costs are typically non-existent, with the key exception being facilitated membrane transport membranes, which require >60% relative humidity to activate the transport mechanism (Ferrari et al, 2021). In this case, the economics of the process are only competitive when waste heat is available within

the plant to raise steam, or when steam can be imported from nearby plants.

Understanding of the cost drivers of each process, particularly the CAPEX to OPEX split, is key to selecting the appropriate technology for CCS projects where public grants and project finance is determinant of investment viability.

4.5 Timeline of CAC developments

The development of CCS technologies implemented within the cement industry over time was also investigated to determine the presence of a 'learning curve'. Due to the availability of data, this analysis was only performed on amine-based and calcium looping technologies.

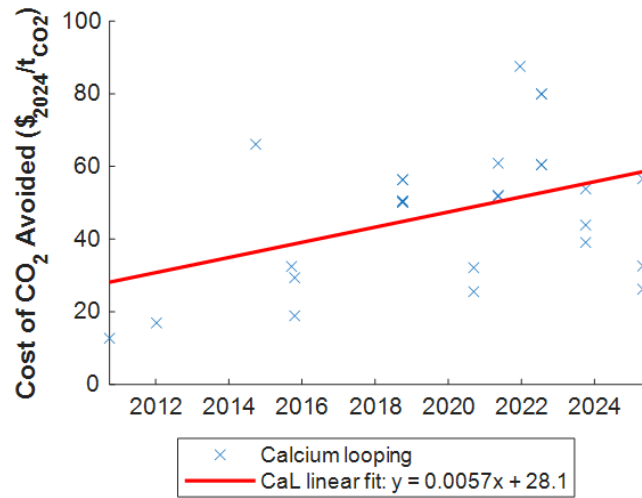


Figure 6: Timeline of Costs of CO₂ Avoided published and linear regression for calcium looping technologies, where x is in days.

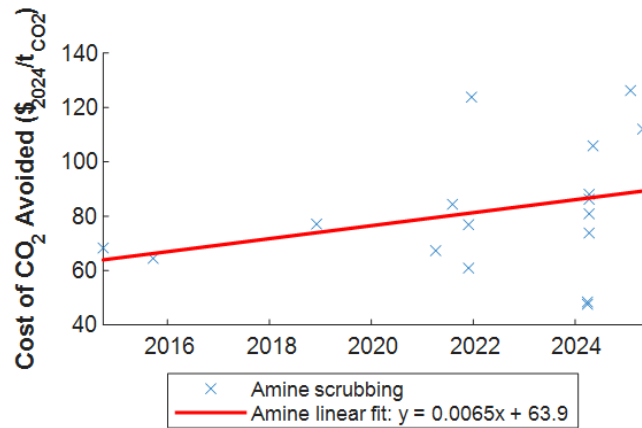


Figure 7: Timeline of Costs of CO₂ Avoided published and linear regression for amine scrubbing technologies, where x is in days

Both Figure 6 and 7 displays the Costs of CO₂ Avoided published in academic literature increasing over time. Earlier studies were performed with simplified assumptions on the process operations, units, and economics, lowering apparent costs. Earlier studies also estimated equipment costs for individual units using a scale factor method and correlations such as Equation 6 below, where C_0 and S_0 represents the reference cost and size, f represents the scale factor, and S_i

as calculated size respectively.

$$C_i = C_0 \left(\frac{S_i}{S_0} \right)^f \quad (6)$$

Academic literature only began to utilise 'bottom-up' methods such as cost functions, Aspen Process Economic Analyser, and other software for CAPEX estimations from 2018, providing more accurate, but more expensive costs.

In an early study by Rodriguez et al, 2012, the authors themselves acknowledge the large uncertainties that exist in the calculated cost due to the scarcity of data providing key capital costs for units such as the precalciner and oxyfired circulating fluidised bed combustion systems.

Earlier papers also assume CaL as mature technology, with later studies applying contingency factors based on the TRL of configurations (De Lena, Spinelli, et al., 2019), subsequently increasing the total plant costs associated. More realistic auxiliary power consumption of units such as the ASU and CPU was also modelled alongside more realistic steam cycle efficiencies.

The earliest study identified for calcium looping from Romeo et al. (2011) combines a coal-fired power plant and cement plant for CO₂ removal, which contributes to the lowest CAC due to the high thermal integration efficiencies compared to a lone cement plant. It is rare for cement plants to be located by a coal power plant, so newer studies incorporate estimates for a standalone boiler to generate steam on site.

4.6 Analysis of capital recovery factor

A range of financial assumptions were observed in the literature reviewed, with discount rates (i) ranging from 6-12%, and operating lifetimes (n) of 15-30 years applied to annualise CAPEX calculations. To ensure comparability of different technologies, the same CRF has been applied in an attempt to harmonise cost reporting. Modal values of an 8% discount rate and an operating lifetime of 25 years have been used to recalculate a CRF of 0.0937, as shown in Equation 7 below.

$$CRF = \frac{i(1+i)^n}{(1+i)^n - 1} \quad (7)$$

$$Ann.CAPEX = TotalPlantCosts \times CRF \quad (8)$$

Figure 8 demonstrates that calcium looping and membrane technologies have the greatest decrease in Cost of Clinker after CRF harmonisation on the initial CAPEX costs. Most literature values apply the same assumptions, but some technoeconomic studies can be seen applying more conservative economic assumptions on the initial CAPEX, either through a higher discount rate, or shorter operating lifetime, or a combination of both. This reflects the lower TRLs observed in industry for calcium looping and membranes, accounting for the inherent risks and uncertainties associated with First-of-a-kind projects.

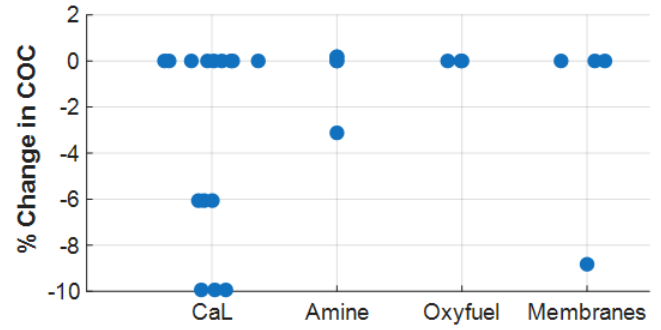


Figure 8: Change in Cost of Clinker after applying the same capital recovery factor to CAPEX in 29 cases.

CAPEX as a % of the Cost of Clinker associated with calcium looping is also higher than other technologies, therefore any change in CRF assumptions would be further amplified. This can be seen with the sensitivity analysis on discount rate performed by Garcia et al. (2018) and Santos and Hanak (2022) where it was observed that CaL's CAC was more sensitive to this parameter compared to oxyfuel combustion and other processes.

However, the TRLs of associated technologies are also accounted for through the different process contingencies built within the Total Plant Costs, with a higher % implying a less mature technology. This can be observed with Gardarsdottir et al. (2019) applying 15% to the more mature MEA scrubbing, 20% to tail-end CaL, and 60% for the fully integrated CaL configurations. Ferrari et al. (2021) applies a 35% process contingency to membranes, compared to the 15% for amine-based separation.

4.7 Policy drivers and commercial investment decisions

To understand the intersection between policy drivers and investment decisions, a comparative analysis has been conducted between the EU ETS and the CAC. If a particular project's CAC is lower than the EU ETS, then it incentivises the cement plant producer/owner to pursue CCS, as it would be more cost-effective than purchasing allowances on the market. However, it is important to note that the EU ETS is a spot price that fluctuates according to market demands, while the CAC is a projection over the plant's operating lifetime of 25-30 years.

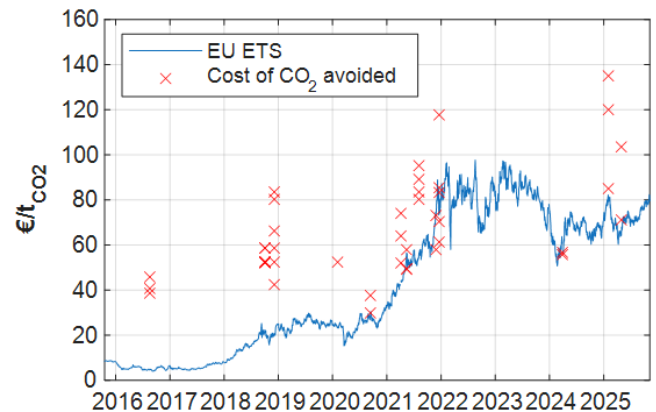


Figure 9: Timeline of Costs of CO₂ Avoided and EU ETS exchange prices. Cost data taken when reported in € and referring to European geographies. (Bloomberg, 2025)

Figure 9 demonstrates the narrowing gap between CAC and EU ETS, with the first 'break-even' project appearing in 2021 due to the spike in EU ETS price. There currently remains a gap between the two, however the EU ETS is forecasted to reach 126 €/tCO₂ by 2030 (Glushchenko, 2015), due to the reduction of free allowances and the full implementation of CBAM in 2026. This has created significant incentives to drive investment in CCS projects across the cement and clinker industry, with public grants provided in some cases to bridge the financing gap and reduce investment risks.

Analysis of 2023 and 2024 annual reports of Europe's 5 largest clinker producers by volume (Heidelberg Materials, Holcim, Cemex, Buzzi, and CRH) also revealed differing attitudes and progress towards CCS in the cement industry. With Cemex, Heidelberg Materials, and CRH explicitly discussing the use of an internal carbon price to guide investment decisions related to CCS projects. Cemex appears to use the most conservative estimate of 145 €/tCO₂ for 2030, and also implements a 'carbon floor price' of 36 USD/tCO₂ in non-regulation regions the company operates in (Cemex, 2024).

Of the 23 pilot or industrial scale CCS projects identified in the annual reports, 19 have confirmed public grants through either national or EU level funding, with all projects bar one expected to be operational by 2030. Despite the abundance of academic literature focused on calcium looping technologies, and the lower relative CAC and SPECCA observed, there are only 2 pilot projects identified utilising CaL. 65% of the projects utilise amine-based scrubbing, oxyfuel, or a hybrid oxyfuel-amine technology, representing a disjunction between results observed in academia and industry.

This can be explained by the technical know-how already present for amine scrubbing and oxyfuel combustion processes in other industries. Amine scrubbing, or gas sweetening, has been utilised in the oil and gas industry since the 1930s to remove CO₂ and H₂S from natural gas streams to meet product specifications and reduce pipe corrosion (Gary, 2007). Oxyfuel combustion has been researched and adopted by the steel industry since the 1980s, with the aim to reduce fuel consumption and increasing steel production output by creating an oxygen-rich furnace (Von, 2010). The decades of operational experience in other industries have allowed deeper research and understanding of topics such as solvent degradation, reaction kinetics, and heat transfer profiles that can be applied to CCS in the cement industry, accelerating the deployment of these technologies and reducing the perceived investment risk.

Although calcium looping appears to be a competitive choice for CCS for the cement industry when considering CAC and SPECCA, the process still lags behind in industry adoption. Despite a successful pilot with the CLEANKER project demonstrating TRL 7, there are still more demonstrations required to understand the attrition and elutriation within scaled-up entrained flow reactors before commercial

deployment at scale (Castillo et al., 2024).

5. Conclusion and future works

The results of this review demonstrate the evolving techno-economic studies of carbon capture systems applied to cement over time, with a focus on oxyfuel combustion, calcium looping, amine scrubbing, and membrane separation technologies. This analysis is primarily conducted through inspection of the Cost of CO₂ Avoided, Specific Primary Energy Consumption for CO₂ Avoided, and Cost of Clinker observed through literature. The authors identified 25 of 138 relevant papers reporting cost data, with the Cost of CO₂ being the most frequently reported cost metric.

Calcium looping emerged as the most economically viable option when considering the Cost of CO₂ Avoided, with the lowest mean of 50.4 \$₂₀₂₄/tCO₂, primarily due to lower net energy demands (surplus electricity exported in some cases). Following this, oxyfuel combustion (65.0 \$₂₀₂₄/tCO₂), membrane separation (78.8 \$₂₀₂₄/tCO₂), and finally amine scrubbing (79.1 \$₂₀₂₄/tCO₂). However, the CAC is highly dependent on economic assumptions, as demonstrated in the review.

Analysis of SPECCA allows for understanding of the energy penalty and indirect emissions associated with each capture system by stripping out economic assumptions that may influence CAC. Oxyfuel reports the lowest SPECCA, the source of the energy penalty being only electricity to power the ASU and CPU, with no regeneration of solvent/sorbent. Calcium looping reports the next lowest SPECCA due to the integration of high grade heat, reducing net energy demands. Membranes exhibit the next highest SPECCA due to electricity demands for compressions/vacuums. Amine scrubbing reports the highest SPECCA due to the energy requirements related to the regeneration of solvent in the stripper.

The authors conclude that the drivers of Cost of CO₂ Avoided are the net energy demands of each technology, and the Specific Primary Energy Consumption for CO₂ Avoided aids understanding the associated energy penalty of each system. As such, it is recommended that future works more frequently report this metric within techno-economic studies, alongside greater standardisation of economic assumptions. More research should also focus on oxyfuel combustion and membrane separations applied to clinker production, to bridge the apparent bias within academic studies.

To bridge the CCS investment gap within the cement industry, greater policy and economic incentives are required. Policy mechanisms such as the CBAM and ETS in the UK and EU have begun to influence commercial investments, but greater adoption is required across more regions to encourage implementation of carbon capture in the global cement industry.

References

Arias, A. M., N. J. Scenna, and P. L. Mores (Oct. 1, 2023). "Membrane superstructure optimization for carbon capture from cement plants. Water content influence on optimal solution". In: *International Journal of Greenhouse Gas Control* 129, p. 103964. ISSN: 1750-5836. DOI: 10.1016/j.ijggc.2023.103964. URL:

- <https://www.sciencedirect.com/science/article/pii/S1750583623001342> (visited on 12/10/2025).
- Aromada, S. A., N. H. Eldrup, and L. E. Øi (Jan. 2022). “Cost and Emissions Reduction in CO₂ Capture Plant Dependent on Heat Exchanger Type and Different Process Configurations: Optimum Temperature Approach Analysis”. In: **Energies** 15.2. Publisher: Multidisciplinary Digital Publishing Institute, p. 425. ISSN: 1996-1073. DOI: 10.3390/en15020425. URL: <https://www.mdpi.com/1996-1073/15/2/425> (visited on 12/10/2025).
- Atsonios, K., P. Grammelis, S. K. Antiohos, N. Nikolopoulos, and E. Kakaras (Aug. 1, 2015). “Integration of calcium looping technology in existing cement plant for CO₂ capture: Process modeling and technical considerations”. In: **Fuel** 153, pp. 210–223. ISSN: 0016-2361. DOI: 10.1016/j.fuel.2015.02.084. URL: <https://www.sciencedirect.com/science/article/pii/S001623611500246X> (visited on 12/07/2025).
- Bacatelo, M., F. Capucha, P. Ferrão, and F. Margarido (Feb. 1, 2023). “Selection of a CO₂ capture technology for the cement industry: An integrated TEA and LCA methodological framework”. In: **Journal of CO₂ Utilization** 68, p. 102375. ISSN: 2212-9820. DOI: 10.1016/j.jcou.2022.102375. URL: <https://www.sciencedirect.com/science/article/pii/S2212982022004942> (visited on 12/10/2025).
- Barbhuiya, S., F. Kanavaris, B. B. Das, and M. Idrees (June 1, 2024). “Decarbonising cement and concrete production: Strategies, challenges and pathways for sustainable development”. In: **Journal of Building Engineering** 86, p. 108861. ISSN: 2352-7102. DOI: 10.1016/j.jobe.2024.108861. URL: <https://www.sciencedirect.com/science/article/pii/S2352710224004297> (visited on 12/10/2025).
- Bloomberg (2025). **EU Carbon Allowance EUA Physically Settled Data (2015-2025)**. (Visited on 11/11/2025).
- Calin-Cristian, C. (July 15, 2022). “Decarbonization options for cement production process: A techno-economic and environmental evaluation”. In: **Fuel** 320, p. 123907. ISSN: 0016-2361. DOI: 10.1016/j.fuel.2022.123907. URL: <https://www.sciencedirect.com/science/article/pii/S0016236122007669> (visited on 12/09/2025).
- Castillo, J. A., B. Wilhelmsson, M. Broström, and M. Eriksson (2024). “Properties of cement raw meals used as sorbents in a calcium looping process”. In: **Chemical Engineering Journal** 499, p. 156165. ISSN: 1385-8947. DOI: <https://doi.org/10.1016/j.cej.2024.156165>. URL: <https://www.sciencedirect.com/science/article/pii/S1385894724076563>.
- CEMBUREAU (2024). **CEMBUREAU Activity Report 2024**. URL: <https://cem bureau.eu/media/aj51xhob/cembureau-activity-report-2024.pdf> (visited on 11/11/2025).
- Cemex (2023). **Cemex 2023 Integrated Report**. URL: <https://www.cemex.com/documents/d/cemex/cemex-2023-integrated-report-en>.
- (2024). **Cemex 2024 Integrated Report**. URL: <https://www.cemex.com/documents/d/cemex/cemex-2024-integrated-report-en>.
- CEPCI (Nov. 26, 2014). **CEPCI Database**. URL: <https://www.cemengonline.com/pci/> (visited on 12/10/2025).
- Chapman, M. (Nov. 13, 2025). **UK Emissions Trading Scheme (ETS) — TotalEnergies Gas & Power**. URL: <https://business.totalenergies.uk/uk-emissions-trading-scheme-ets> (visited on 12/11/2025).
- Chauvy, R., D. Verdonck, L. Dubois, D. Thomas, and G. De Weireld (May 1, 2021). “Techno-economic feasibility and sustainability of an integrated carbon capture and conversion process to synthetic natural gas”. In: **Journal of CO₂ Utilization** 47, p. 101488. ISSN: 2212-9820. DOI: 10.1016/j.jcou.2021.101488. URL: <https://www.sciencedirect.com/science/article/pii/S221298202100055X> (visited on 12/10/2025).
- Cloete, S., A. Giuffrida, M. C. Romano, and A. Zaabout (Dec. 1, 2020). “Economic assessment of the swing adsorption reactor cluster for CO₂ capture from cement production”. In: **Journal of Cleaner Production** 275, p. 123024. ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2020.123024. URL: <https://www.sciencedirect.com/science/article/pii/S0959652620330699> (visited on 12/10/2025).
- Cormos, C.-C. (July 15, 2022). “Decarbonization options for cement production process: A techno-economic and environmental evaluation”. In: **Fuel** 320, p. 123907. ISSN: 0016-2361. DOI: 10.1016/j.fuel.2022.123907. URL: <https://www.sciencedirect.com/science/article/pii/S0016236122007669> (visited on 12/10/2025).
- Cormos, C. and Cormos A. (Oct. 2015). “Assessment of CO₂ capture by calcium looping from natural gas combined cycle (ngcc) power plants”. In: **Chemical Engineering Transactions** 45, pp. 277–282. DOI: 10.3303/CET1545047. URL: <https://doi.org/10.3303/CET1545047> (visited on 12/07/2025).
- Dashtizadeh, E. and E. Houshfar (Nov. 4, 2024). “Comparative thermoeconomic analysis of integrated hybrid multigeneration systems with hydrogen production for waste heat recovery in cement plants”. In: **International Journal of Hydrogen Energy** 89, pp. 868–882. ISSN: 0360-3199. DOI: 10.1016/j.ijhydene.2024.09.359. URL: <https://www.sciencedirect.com/science/article/pii/S0360319924040849> (visited on 12/10/2025).
- De Lena, E., B. Arias, M. C. Romano, and J. C. Abanades (Feb. 1, 2022). “Integrated Calcium Looping System with Circulating Fluidized Bed Reactors for Low CO₂ Emission Cement Plants”. In: **International Journal of Greenhouse Gas Control** 114, p. 103555. ISSN: 1750-5836. DOI: 10.1016/j.ijggc.2021.103555. URL: <https://www.sciencedirect.com/science/article/pii/S1750583621003066> (visited on 12/10/2025).
- De Lena, E., M. Spinelli, M. Gatti, R. Scaccabarozzi, S. Campanari, S. Consonni, G. Cinti, and M. C. Romano (Mar. 1, 2019). “Techno-economic analysis of calcium looping processes for low CO₂ emission cement plants”. In: **International Journal of Greenhouse Gas Control** 82, pp. 244–260. ISSN: 1750-5836. DOI: 10.1016/j.ijggc.2019.01.005. URL: <https://www.sciencedirect.com/science/article/pii/S1750583618307473> (visited on 12/10/2025).
- Diego, M. E., B. Arias, and J. C. Abanades (Mar. 20, 2016). “Analysis of a double calcium loop process configuration for CO₂ capture in cement plants”. In: **Journal of Cleaner Production** 117, pp. 110–121. ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2016.01.027. URL: <https://www.sciencedirect.com/science/article/pii/S0959652616000536> (visited on 12/10/2025).
- European Commission, J. R. C. (2023). **Decarbonisation options for the cement industry**. Publications Office of the European Union. DOI: [doi/10.2760/174037](https://doi.org/10.2760/174037).
- Faria, D. G., M. M. O. Carvalho, M. R. V. Neto, E. C. de Paula, M. Cardoso, and E. K. Vakkilainen (Feb. 1, 2022). “Integrating oxy-fuel combustion and power-to-gas in the cement industry: A process modeling and simulation study”. In: **International Journal of Greenhouse Gas Control** 114, p. 103602. ISSN: 1750-5836. DOI: 10.1016/j.ijggc.2022.103602. URL: <https://www.sciencedirect.com/science/article/pii/S1750583622000226> (visited on 12/10/2025).

- Ferrari, M.-C., A. Amelio, G. M. Nardelli, and R. Costi (Jan. 2021). "Assessment on the Application of Facilitated Transport Membranes in Cement Plants for CO₂ Capture". In: **Energies** 14.16. Publisher: Multidisciplinary Digital Publishing Institute, p. 4772. ISSN: 1996-1073. DOI: 10.3390/en14164772. URL: <https://www.mdpi.com/1996-1073/14/16/4772> (visited on 12/10/2025).
- Ferrario, D., T. Pröll, S. Stendardo, and A. Lanzini (Aug. 15, 2024). "Cost and environmentally efficient design of an absorption-based post-combustion carbon capture unit for industry applications". In: **Chemical Engineering Journal** 494, p. 152900. ISSN: 1385-8947. DOI: 10.1016/j.cej.2024.152900. URL: <https://www.sciencedirect.com/science/article/pii/S1385894724043870> (visited on 12/10/2025).
- Garcia, M., N. Berghout, and L. Herraiz (Sept. 4, 2018). **Co₂ Capture in the Cement and Iron and Steel Sectors: A Homogenised Techno-Economic Review**. Rochester, NY. DOI: 10.2139/ssrn.3366376. URL: <https://papers.ssrn.com/abstract=3366376> (visited on 12/08/2025).
- Gardarsdottir, S. O., E. De Lena, M. Romano, S. Roussanaly, M. Voldsund, J.-F. Pérez-Calvo, D. Berstad, C. Fu, R. Anantharaman, D. Sutter, M. Gazzani, M. Mazzotti, and G. Cinti (Jan. 2019). "Comparison of Technologies for CO₂ Capture from Cement Production—Part 2: Cost Analysis". In: **Energies** 12.3. Publisher: Multidisciplinary Digital Publishing Institute, p. 542. ISSN: 1996-1073. DOI: 10.3390/en12030542. URL: <https://www.mdpi.com/1996-1073/12/3/542> (visited on 12/10/2025).
- Gary, J. H. (2007). **Petroleum refining: technology and economics**. In collab. with G. E. Handwerk and M. J. Kaiser. 5th ed. Boca Raton: CRC Press. xix+465. ISBN: 978-0-8493-7038-0.
- Glushchenko, A. (Dec. 2, 2015). **Carbon price in the EU ETS to hit €126/t by 2030**. GMK. URL: <https://gmk.center/en/infographic/carbon-price-in-the-eu-ets-to-hit-e126-t-by-2030/> (visited on 12/07/2025).
- Hägg, M. -, A. Lindbråthen, X. He, S. G. Nodeland, and T. Cantero (July 1, 2017). "Pilot Demonstration-reporting on CO₂ Capture from a Cement Plant Using Hollow Fiber Process". In: **Energy Procedia**. 13th International Conference on Greenhouse Gas Control Technologies, GHGT-13, 14-18 November 2016, Lausanne, Switzerland 114, pp. 6150–6165. ISSN: 1876-6102. DOI: 10.1016/j.egypro.2017.03.1752. URL: <https://www.sciencedirect.com/science/article/pii/S1876610217319549> (visited on 12/10/2025).
- Hughes, S., P. Cvetic, R. Newby, S. Homsy, A. Zoelle, M. Woods, E. Grol, and T. Fout (June 1, 2024). "The Impact of Cement Plant Air Ingress on Membrane-Based CO₂ Capture Retrofit Cost". In: **Carbon Capture Science & Technology** 11, p. 100201. ISSN: 2772-6568. DOI: 10.1016/j.ccst.2024.100201. URL: <https://www.sciencedirect.com/science/article/pii/S2772656824000137> (visited on 12/10/2025).
- IEA (2025). **Global Energy and Climate Model**. IEA. URL: <https://www.iea.org/reports/global-energy-and-climate-model/net-zero-emissions-by-2050-scenario-nze> (visited on 12/10/2025).
- IMF (2025). **IMF Data Explorer**. URL: <https://data.imf.org/en/Data-Explorer?> (visited on 12/10/2025).
- Jiang, K., H. Yu, Z. Sun, Z. Lei, K. Li, and L. Wang (Apr. 23, 2024). "Zero-Emission Cement Plants with Advanced Amine-Based CO₂ Capture". In: **Environmental Science & Technology** 58.16. Publisher: American Chemical Society, pp. 6978–6987. ISSN: 0013-936X. DOI: 10.1021/acs.est.4c00197. URL: <https://doi.org/10.1021/acs.est.4c00197> (visited on 12/10/2025).
- Kvamsdal, H. M., G. Haugen, and N. H. Eldrup (Apr. 6, 2021). **Benchmarking of a Membrane Technology for Tail-end Application in a Cement Plant**. Rochester, NY. DOI: 10.2139/ssrn.3820500. URL: <https://papers.ssrn.com/abstract=3820500> (visited on 12/10/2025).
- Lee, S. and J.-K. Kim (Nov. 15, 2023). "Sub-ambient membrane process for CO₂ removal in the industrial sector: Iron and steel, cement, and refinery". In: **Journal of Membrane Science** 686, p. 122018. ISSN: 0376-7388. DOI: 10.1016/j.memsci.2023.122018. URL: <https://www.sciencedirect.com/science/article/pii/S0376738823006749> (visited on 12/10/2025).
- Leeson, D., N. Mac Dowell, N. Shah, C. Petit, and P. Fennell (June 2017). "A Techno-economic analysis and systematic review of carbon capture and storage (CCS) applied to the iron and steel, cement, oil refining and pulp and paper industries, as well as other high purity sources". In: **International Journal of Greenhouse Gas Control** 61, pp. 71–84. ISSN: 17505836. DOI: 10.1016/j.ijggc.2017.03.020. URL: <https://linkinghub.elsevier.com/retrieve/pii/S175058361730289X> (visited on 12/10/2025).
- Liu, X., K. Jin, X. Li, and R. Yang (Oct. 1, 2023). "Low-carbon cement manufacturing enabled by electrified calcium looping and thermal energy storage". In: **International Journal of Greenhouse Gas Control** 129, p. 103986. ISSN: 1750-5836. DOI: 10.1016/j.ijggc.2023.103986. URL: <https://www.sciencedirect.com/science/article/pii/S1750583623001561> (visited on 12/10/2025).
- Magli, F., M. Spinelli, M. Fantini, M. C. Romano, and M. Gatti (Mar. 1, 2022). "Techno-economic optimization and off-design analysis of CO₂ purification units for cement plants with oxyfuel-based CO₂ capture". In: **International Journal of Greenhouse Gas Control** 115, p. 103591. ISSN: 1750-5836. DOI: 10.1016/j.ijggc.2022.103591. URL: <https://www.sciencedirect.com/science/article/pii/S1750583622000135> (visited on 12/10/2025).
- Martin, A. C., F. Lacouture, P. Llewellyn, and L. Mariac (Dec. 9, 2021). "A Fair Comparison Between Five CO₂ Capture Technologies". In: Abu Dhabi International Petroleum Exhibition & Conference. OnePetro. DOI: 10.2118/208138-MS. URL: <https://dx.doi.org/10.2118/208138-MS> (visited on 12/10/2025).
- Micari, M., M. Dakhchoune, and K. V. Agrawal (Apr. 15, 2021). "Techno-economic assessment of postcombustion carbon capture using high-performance nanoporous single-layer graphene membranes". In: **Journal of Membrane Science** 624, p. 119103. ISSN: 0376-7388. DOI: 10.1016/j.memsci.2021.119103. URL: <https://www.sciencedirect.com/science/article/pii/S0376738821000569> (visited on 12/10/2025).
- Michailos, S., M. Lucquiaud, A. D. Benz, D. Mullen, W. Elliott, and J. Gibbins (Apr. 2, 2025). "A Performance Modeling Study for Zero Fossil CO₂ Stack Operation and Solvent Thermal Reclaiming in Post-Combustion Capture Industrial Applications". In: **Industrial & Engineering Chemistry Research** 64.13. Publisher: American Chemical Society, pp. 7100–7114. ISSN: 0888-5885. DOI: 10.1021/acs.iecr.4c04530. URL: <https://doi.org/10.1021/acs.iecr.4c04530> (visited on 12/10/2025).
- Nelson, T. O., A. Kataria, P. Mobley, M. Soukri, and J. Tanthana (July 1, 2017). "RTI's Solid Sorbent-Based CO₂ Capture Process: Technical and Economic Lessons Learned for Application in Coal-fired, NGCC, and Cement Plants". In: **Energy Procedia**. 13th International Conference on Greenhouse Gas Control Tech-

- nologies, GHGT-13, 14-18 November 2016, Lausanne, Switzerland 114, pp. 2506–2524. ISSN: 1876-6102. doi: 10.1016/j.egypro.2017.03.1409. URL: <https://www.sciencedirect.com/science/article/pii/S1876610217315941> (visited on 12/10/2025).
- Nwaoha, C., M. Beaulieu, P. Tontiwachwuthikul, and M. D. Gibson (Nov. 1, 2018). “Techno-economic analysis of CO₂ capture from a 1.2 million MTPA cement plant using AMP-PZ-MEA blend”. In: **International Journal of Greenhouse Gas Control** 78, pp. 400–412. ISSN: 1750-5836. doi: 10.1016/j.ijggc.2018.07.015. URL: <https://www.sciencedirect.com/science/article/pii/S1750583618303566> (visited on 12/10/2025).
- Rezvani, S., A. Rolfe, F. Franco, C. Brandoni, K. Böge, N. Hewitt, and Y. Huang (Mar. 15, 2025). “Indirectly heated carbonate looping cycles in cement plants for CO₂ capture and storage”. In: **Applied Thermal Engineering** 263, p. 125349. ISSN: 1359-4311. doi: 10.1016/j.applthermaleng.2024.125349. URL: <https://www.sciencedirect.com/science/article/pii/S1359431124030175> (visited on 12/10/2025).
- Rodríguez, N., R. Murillo, and J. C. Abanades (Feb. 21, 2012a). “CO₂ Capture from Cement Plants Using Oxyfired Precalcination and/or Calcium Looping”. In: **Environmental Science & Technology** 46.4. Publisher: American Chemical Society, pp. 2460–2466. ISSN: 0013-936X. doi: 10.1021/es2030593. URL: <https://doi.org/10.1021/es2030593> (visited on 12/07/2025).
- (Feb. 21, 2012b). “CO₂ Capture from Cement Plants Using Oxyfired Precalcination and/or Calcium Looping”. In: **Environmental Science & Technology** 46.4. Publisher: American Chemical Society, pp. 2460–2466. ISSN: 0013-936X. doi: 10.1021/es2030593. URL: <https://doi.org/10.1021/es2030593> (visited on 12/10/2025).
- Rolfe, A., Y. Huang, M. Haaf, A. Pita, S. Rezvani, A. Dave, and N. J. Hewitt (Aug. 1, 2018). “Technical and environmental study of calcium carbonate looping versus oxy-fuel options for low CO₂ emission cement plants”. In: **International Journal of Greenhouse Gas Control** 75, pp. 85–97. ISSN: 1750-5836. doi: 10.1016/j.ijggc.2018.05.020. URL: <https://www.sciencedirect.com/science/article/pii/S1750583617309775> (visited on 12/10/2025).
- Romeo, L. M., D. Catalina, P. Lisbona, Y. Lara, and A. Martínez (2011). “Reduction of greenhouse gas emissions by integration of cement plants, power plants, and CO₂ capture systems”. In: **Greenhouse Gases: Science and Technology** 1.1. eprint: <https://scijournals.onlinelibrary.wiley.com/doi/pdf/10.1002/ghg3.5>, pp. 72–82. ISSN: 2152-3878. doi: 10.1002/ghg3.5. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ghg3.5> (visited on 12/07/2025).
- Santos, M. P. S. and D. P. Hanak (Sept. 1, 2022). “Carbon capture for decarbonisation of energy-intensive industries: a comparative review of techno-economic feasibility of solid looping cycles”. In: **Frontiers of Chemical Science and Engineering** 16.9, pp. 1291–1317. ISSN: 2095-0187. doi: 10.1007/s11705-022-2151-5. URL: <https://doi.org/10.1007/s11705-022-2151-5> (visited on 12/08/2025).
- Scholes, C. A., M. T. Ho, A. A. Aguiar, D. E. Wiley, G. W. Stevens, and S. E. Kentish (May 1, 2014). “Membrane gas separation processes for CO₂ capture from cement kiln flue gas”. In: **International Journal of Greenhouse Gas Control** 24, pp. 78–86. ISSN: 1750-5836. doi: 10.1016/j.ijggc.2014.02.020. URL: <https://www.sciencedirect.com/science/article/pii/S1750583614000589> (visited on 12/10/2025).
- Shamsi, M., E. Naeiji, M. Vaziri, S. Moghaddas, K. Elyasi Gommari, M. Naseri, and M. Bonyadi (Apr. 1, 2025). “Optimization of energy requirements and 4E analysis for an integrated post-combustion CO₂ capture utilizing MEA solvent from a cement plant flue gas”. In: **Environment, Development and Sustainability** 27.4, pp. 9379–9405. ISSN: 1573-2975. doi: 10.1007/s10668-023-04287-6. URL: <https://doi.org/10.1007/s10668-023-04287-6> (visited on 12/10/2025).
- Strojny, M., P. Suchodolski, P. Gładysz, and W. Nowak (July 1, 2025). “Preliminary analysis of combining amine absorption and calcium-looping CO₂ capture in cement plant”. In: **Energy** 327, p. 136361. ISSN: 0360-5442. doi: 10.1016/j.energy.2025.136361. URL: <https://www.sciencedirect.com/science/article/pii/S0360544225020031> (visited on 12/07/2025).
- Svaestuen, J. (2017). **CO₂ Capture Test Facility at Norcem Brevik, Project Phase II, Test Step 1**. Climit. URL: <https://climit.no/en/project/co2-capture-test-facility-at-norcem-brevik-project-phase-ii-test-step-1/> (visited on 12/11/2025).
- Turakulov, Z., A. Kamolov, A. Norkobilov, M. Variny, and M. Faltanza (Oct. 1, 2024). “Techno-economic and environmental analysis of decarbonization pathways for cement plants in Uzbekistan”. In: **Chemical Engineering Research and Design** 210, pp. 625–637. ISSN: 0263-8762. doi: 10.1016/j.cherd.2024.09.003. URL: <https://www.sciencedirect.com/science/article/pii/S026387622400532X> (visited on 12/10/2025).
- United Nations, E. P. (Nov. 2025). **Emissions Gap Report 2025: Off Target - Continued Collective inaction puts Global Temperature Goal at Risk**. ISBN: 978-92-807-4239-8. doi: 10.59117/20.500.11822/48854. URL: <https://wedocs.unep.org/handle/20.500.11822/48854> (visited on 12/10/2025).
- Varnier, L., F. d’Amore, K. Clausen, G. Melitos, B. de Groot, and F. Bezzo (Mar. 20, 2025). “Combined electrification and carbon capture for low-carbon cement: Techno-economic assessment of different designs”. In: **Journal of Cleaner Production** 498, p. 145029. ISSN: 0959-6526. doi: 10.1016/j.jclepro.2025.145029. URL: <https://www.sciencedirect.com/science/article/pii/S0959652625003798> (visited on 12/10/2025).
- Von, J. (Aug. 17, 2010). “Oxyfuel Combustion in the Steel Industry: Energy Efficiency and Decrease of CO₂ Emissions”. In: **Energy Efficiency**. Ed. by J. Palm. Sciyo. ISBN: 978-953-307-137-4. doi: 10.5772/9830. URL: <http://www.intechopen.com/books/energy-efficiency/oxyfuel-combustion-in-the-steel-industry-energy-efficiency-and-decrease-of-co2-emissions> (visited on 12/08/2025).
- Williams, F., A. Yang, and D. R. Nhuchhen (Jan. 15, 2024). “Decarbonisation pathways of the cement production process via hydrogen and oxy-combustion”. In: **Energy Conversion and Management** 300, p. 117931. ISSN: 0196-8904. doi: 10.1016/j.enconman.2023.117931. URL: <https://www.sciencedirect.com/science/article/pii/S0196890423012773> (visited on 12/10/2025).
- Wu, W., Z.-L. Jian, B.-Y. Chou, C.-Y. You, and Y.-N. Kuo (July 2023). “Process Design and Techno-ECONOMIC Evaluation of a Decarbonized Cement Production Process Using Carbon Capture and Utilization”. In: **Processes** 11.7. Publisher: Multidisciplinary Digital Publishing Institute, p. 2043. ISSN: 2227-9717. doi: 10.3390/pr11072043. URL: <https://www.mdpi.com/2227-9717/11/7/2043> (visited on 12/10/2025).
- Yin, J., C. Li, G. Paicu, and S. Su (Mar. 1, 2024). “Techno-economic assessment of retrofitting indirect-heated calcium looping using coal and biomass as fuels into an existing cement

plant for CO₂ capture”. In: **Gas Science and Engineering** 123, p. 205236. ISSN: 2949-9089. DOI: 10.1016/j.jgsce.2024.205236. URL: <https://www.sciencedirect.com/science/article/pii/S2949908924000323> (visited on 12/10/2025).

Zanobetti, F., A. Dal Pozzo, and V. Cozzani (Feb. 1, 2025). “Sustainability assessment of CO₂ capture across different scales of hard-to-abate emission sources”. In: **Chemical Engineering Journal** 505, p. 159466. ISSN: 1385-8947. DOI: 10.1016/j.cej.2025.159466. URL: <https://www.sciencedirect.com/science/article/pii/S1385894725002657> (visited on 12/10/2025).

Zheng, Y., L. Gao, S. He, X. Zeng, J. Wang, J. Liu, W. Wang, and B. Xu (Nov. 1, 2025). “Technical, economic and environmental assessments of integrating calcium looping and chemical looping hydrogen generation processes for existing cement plants decarbonization”. In: **Process Safety and Environmental Protection** 203, p. 107711. ISSN: 0957-5820. DOI: 10.1016/j.psep.2025.107711. URL: <https://www.sciencedirect.com/science/article/pii/S0957582025009784> (visited on 12/10/2025).

Evaluating Strategies to Reduce Experimental Load for Design Space Identification

Idrees Kholwadia^{b,1} and Kundan Mahitkumar^{a,2}

^aSupervisor: Maria Papathanasiou

^bDepartment of Chemical Engineering

^cImperial College London

Abstract—Design space identification is necessary to give operators insight on optimising the inputs, efficiency and efficacy of their process. However, this is hindered due to computationally heavy mechanistic models or scarce availability of experimental data, as is the case in most bioprocesses and periodic systems. Hence, this paper proposes a data-driven framework to mitigate these challenges by integrating neural network models in a Multicolumn Counter-current Solvent Gradient Purification (MCSGP) system. Firstly, artificial neural networks (ANN) were trained across various sample sizes to serve as surrogate models to estimate KPIs and generate a precise design space. The ANN model trained on 600 samples was deemed optimal, achieving a high Jaccard Index of 0.656 and 79.7% accuracy, demonstrating a crucial balance between computational efficiency and high-fidelity design space identification. Then, ANNs were implemented within the Optimisation and Machine Learning Toolkit (OMLT) to reformulate the system as Mixed-Integer Linear Program (MILP), enabling the reverse-solution of optimal operating inputs. Secondly, Bayesian neural networks (BNN) were implemented where their inherent probabilistic nature characterised variables by quantifying the probability of satisfying the set KPI constraints, providing operators with critical confidence bounds for operation. The probabilistic design space generated from the BNN showed a lower Jaccard Index value than the original design space. Nevertheless, it provided substantial qualitative insight that enhanced overall process understanding. This paper validates the integration of neural networks to reduce experimental effort and elucidate decision-making in complex periodic systems.

Keywords—machine learning, neural networks, design space, process optimisation, experimental efficiency

1. Introduction

Many bioprocesses are highly sensitive to operating conditions and have strict requirements for key performance indicators (KPI). Experimental trials to characterise these systems are typically expensive and time-consuming, limiting the combination of input variables that can be explored. This means that often times the operating point is far from the optimum. As a result, any method that can reduce the cost of evaluating different input conditions is extremely valuable [1].

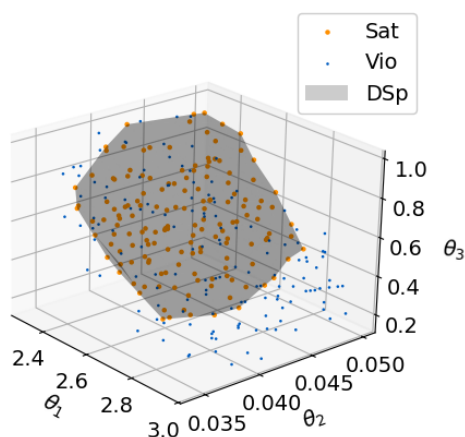


Figure 1. Generic design space, shown by the black shape. It is defined by set constraints, where θ represents the input parameters.

One of the existing strategies to solve this is design space identification[2]. A design space (Figure 1) is a visual representation of all the feasible operating points

that will satisfy the KPI constraints. It is generated by running experimental trials on a range of up to 3 input variables and obtaining their corresponding KPI values. Then inputs that satisfy performance requirements are grouped together using alpha shapes algorithms. This process defines a region within the input space that represents the acceptable operating conditions, without having to experimentally validate each point in the region. By selecting a point within the centre of the design space operators can ensure flexibility.

Although design spaces can reduce the cost of identifying an operating range, constructing a well-informed space still requires a large number of experimental trials. A common strategy to lower this burden is to train neural-network-based surrogate models on the available experimental data and use them to interpolate the unexplored regions. Additionally, probabilistic design spaces, where feasible regions in the input space are expressed with associated confidence levels, can be constructed using machine-learning models that produce predictive distributions. Chiplunkar, et al [3] achieved this using GPs for an antibody-producing Chinese hamster ovary (CHO) cell culture process.

Machine Learning techniques, have been widely applied as surrogates in bioprocess modelling, optimisation, and control [4]. However, despite the extensive use of machine-learning models in bioprocess development, their application to design-space construction has been far less critically assessed. Existing studies typically focus on predicting process outputs or improving model accuracy, but do not evaluate how well the resulting surrogate-derived design spaces represent the true feasible region.

This gap highlights an important observation: model accuracy and design-space accuracy cannot be assumed to be equivalent.

In this paper, Artificial Neural Networks (ANNs) and Bayesian Neural Networks (BNNs) are investigated as methods to improve design space accuracy, in the context of the MCSGP process (see section 2). Section 3 provides a framework to evaluate and compare design spaces. Section 4 then follows by demonstrating how NNs can significantly improve design space accuracy and process understanding. Finally, section 5 concludes by comparing both approaches and commenting on which has performed better.

2. Background

2.1. MCSGP Process

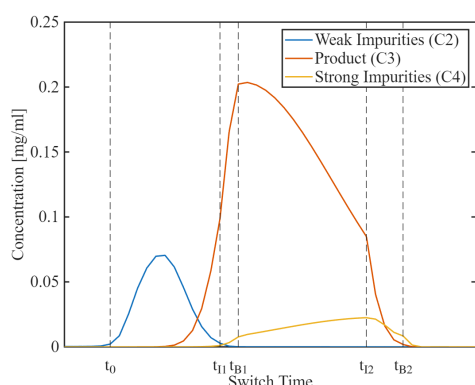


Figure 2. Elution profile over one cycle, where t_0 is start time, t_{I1} is interconnected phase 1, t_{B1} is batch phase 1, t_{I2} is interconnected phase 2, and t_{B2} is batch phase 2.

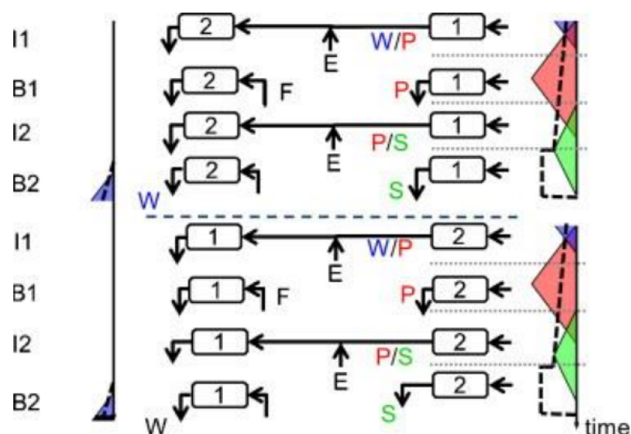


Figure 3. MCSGP process diagram representing the two columns (1 and 2). The elution gradient profile shown on the right. W is weak impurities, P is product, S is strong impurities, E is eluent and F is feed.[5]

The design space identification was carried out on a Multicolumn Counter-current Solvent Gradient Purification process. Initially developed by Aumann and Morbidelli [6], the process utilises ion-exchange chromatography for the purification of biomolecules. The MCSGP process is a semi-continuous mode of chromatography,

consisting of two identical columns whose roles are periodically switched. MCSGP is utilised when the product of interest is embedded in a mixture with multiple impurities, of both shorter and longer elution time (see Figure 2). The basic idea behind the different phases of the process is that, whenever a peak overlap is eluting, the system operates sends material to the second column to be recycled. Conversely, when a pure component is eluting, the system switches phase and directs the stream to its intended destination. The recycling of overlapping peaks allows MCSGP to achieve much greater yields than batch chromatography without sacrificing purity.

The mixture is defined as having a weakly binding/-fast eluting component (W), a product (P) and a strongly binding/slow eluting component (S). The entire cycle, shown in Figure 3, is split into four stages that alternate between interconnected phases (I1, I2) and batch phases (B1, B2). The cycle begins with phase I1, where the effluent of column one is a mixture of W and P. Since this mixture contains product, it is combined with eluent and sent to the second column to further separate the two components. Once the peak overlap between W and P subsides and pure product is exiting column one, the system transitions to phase B1. Here, the connection between the columns is interrupted, allowing the collection of P (whilst fresh feed is concurrently supplied to column 2). Once the strong impurity begins to elute, the P-S peak overlap begins. This marks the start of phase I2, where the connection to column 2 is reopened allowing for the P and S peak to separate further. Finally, once all of the product has left column one and only strong impurities are eluting, phase B2 begins. The system operates in batch mode to remove the strong impurities from column 1 and the weak impurities from column 2. The cycle concludes when the fraction of weak impurities and product reaches the bottom of column 2, at which point the columns switch roles and phase I1 recommences, ensuring semi-continuous operation.

The description above seems incredibly ideal, in the sense that none of the components of interest entering column two elute until the final stage. Meanwhile, over that same period, all components of interest pass through and elute from column one. This is achieved via maintaining both columns at different solvent gradients. Solvent strength, dictated by the modifier concentration, determines the elution velocity of each component: increasing the solvent strength weakens the analyte-stationary phase interactions and accelerates the elution of all species. Solvent gradients are where the solvent strength of the mixture gradually increases over time. By operating the two columns at staggered positions along their gradient profiles, MCSGP ensures that nothing elutes from column two until the end of the cycle. Given that component velocities are integral to the process, any parameter that influences it is important to optimise.

2.2. High Fidelity Model

A mechanistic model is a mathematical representation of processes, built upon the underlying physiochemical principles that govern them, and evaluated by series of

equations. As such is the high-fidelity model (HFM), developed by Michalopolou[7], which utilises lumped mass transfer kinetics and comprehensive mass balances to model the four phases of the cycle and the complete process. Complexities arose when calculating solid and liquid phase concentrations due to the equations being space and time dependent. This was remediated by using 50 discretization points across the column length. The framework resulted in a large-scale model defined by 3309 variables, 4119 equations, of which 805 were ODEs, necessitating immense computational power to solve [7].

Table 1. Input variables and their descriptions.

Input	Description
CFEED2	Feed concentration of weak impurities
CFEED3	Feed concentration of product
CFEED4	Feed concentration of strong impurities
QB	Feed flow rate
QI1	Inlet flow rate of the column executing the gradient elution during phase I1
QI2	Inlet flow rate of the column executing the gradient elution during phase I2
CFEED1	Feed concentration of modifier
CMOD1	Initial modifier concentration for the column executing the gradient elution
CMOD2	Initial modifier concentration for the column executing the recycling and feeding tasks

The HFM requires the 9 inputs, listed in Table 1, to be specified. After the simulation, the model produced detailed concentrations profiles, purity and yield over 1361 minutes and requires a simulation times of 89 seconds per point [7].

2.3. Cyclic Steady State

Given the inherent periodicity and semi-continuous nature of the MCSGP process, its stable operation is defined by Cyclic Steady State (CSS) operation. Upon reaching CSS, the system can extirpate any input-related disturbances. The defining characteristic of CSS is the superimposition of data, where the concentration profiles of individual components and KPIs become indistinguishable between consecutive cycles, given that the batch time is constant. Consequently, all data points used for modelling and validation are exclusively sourced from the CSS region.

3. Methodology

3.1. Determination of 3 Input Variables

Table 2. Inputs considered for data generation.

Input	Baseline Value	Lower Bound	Upper Bound	±%	Units
CFEED2	0.07	0.05	0.08	20	mg/ml
CFEED3	0.40	0.30	0.40	20	mg/ml
CFEED4	0.04	0.03	0.05	20	mg/ml
QB	1.00	0.50	1.00	–	ml/min
QI1	0.10	0.10	1.00	–	ml/min
QI2	0.10	0.10	1.00	–	ml/min
CFEED1	2.00	1.60	2.40	20	mg/ml
CMOD1	2.50	2.00	3.00	20	mg/ml
CMOD2	1.00	0.80	1.20	20	mg/ml

As previously mentioned, the visual nature of design spaces limits the analysis to 3 dimensions. Consequently, this investigation requires strategically selecting and investigating three variables whilst the remaining 6 variables are held at their baseline values. The input parameters are split into 3 distinct subcategories: modifier concentration (CMOD1, CMOD2, CFEED1), flow rates (QI1, QI2, QB), and feed concentration (CFEED2, CFEED3, CFEED4).

To ensure that the selected variables possess independent effects on the KPIs, a single representative parameter was chosen from each subcategory. This selection was systematically executed by performing an initial sensitivity analysis on all 9 variables. As detailed in Table 2, the analysis isolated each parameter, evaluating its impact on purity and yield across its feasible operating range whilst all other variables were kept constant.

Feed flow rate (QB) was quickly eliminated as manipulative variable due to the minimal impact on the KPIs when varied over the feasible operating range. Its baseline value also occurred to be at the maximum, rendering it redundant for investigative purposes. QI2 was discounted due to its primary influence being limited to the elution of the weak impurities, hence its lacklustre effect on the KPIs relative to QI1. QI1 was pursued as it influences the elution gradient during phase I1 and preparing the column for the subsequent feed task during phase B1. This was expected to yield a greater influence on the KPIs, in comparison to QI2.

A stronger modifier concentration increases the competition for the resin's binding sites, which displaces the impurities and products, resulting in faster elution. Therefore, the elution must have a precisely controlled gradient to separate the components properly, without contaminating the final product. The modifier in the first column (CMOD1) is primarily responsible for creating the gradient elution whilst the modifier entering the second column (CMOD2) executes the recycling and feeding of the solvents. In theory, varying CMOD1 would induce a greater deviation in the KPIs, therefore the reason to pursue it over CMOD2.

CFEED3 (product) was ultimately chosen due to its hypothesised direct and significant correlation with the KPIs. This anticipated strong linkage made CFEED3 an attractive variable for investigating the process compared to the impurities. The final selected set of variables for investigation were CMOD1, CFEED3 and QI1.

3.2. Determination of the 3 KPIs

In addition to selecting the three input variables, it is necessary to identify the KPIs to investigate, as these form the basis of the operator's constraints. The first KPI was purity; the most logical choice as the main reason of the MCSGP process is to purify the product. Meeting a certain threshold is necessary for its efficacy, quality and economic potential.

$$\text{Purity [\%]} = \frac{C_{av,P}}{C_{av,W} + C_{av,P} + C_{av,S}} \times 100 \quad (1)$$

The second KPI is the product yield, an important mea-

sure that quantifies the amount of product retrieved successfully. Yield enables comparison between processes and minimises product loss for both financial and efficiency benefits. Moreover, one of the main advantages of MCSGP is its ability to achieve higher yields through recycling overlapping fractions. Failing to optimise for yield therefore underutilises the potential of the process.

$$\text{Yield [\%]} = \frac{C_{\text{av,P}}}{C_{\text{p}}^{\text{feed}}} \times 100 \quad (2)$$

Although purity and yield capture the majority of the operational performance, the sustainability of the process has not yet been considered. Hence, solvent utilisation is an exciting prospect to understand the ratio of the consumption of solvents (primarily being the modifier) to the mass of product purified.

$$\text{Solvent Utilisation [mL/mg]} = \frac{Q_1 + Q_2}{\text{CFEED3} \times \text{Yield} \times \text{QB}} \quad (3)$$

Because this study is not tied to a specific industrial scenario, there are no predetermined KPI targets. Initial sampling with the deterministic model was therefore used to understand the range of achievable performance, and this information informed the selection of KPI constraints. Since the aim of this work is to assess design space accuracy, rather than optimise a real process, the KPI thresholds were chosen to produce informative and well-defined design spaces rather than to reflect strict operational requirements. On this basis, targets of 95% purity, 85% yield, and 8 mL/mg solvent utilisation were selected for the constrained design space.

3.3. Generating Design Spaces

Large sets of input data were generated for the simulations using Sobol sequencing, a low discrepancy quasi-random sequence that ensures the coverage of the whole feasible operating region for each variable. Using MATLAB, they were inputted to the HFM in gPROMS ModelBuilder v7.1.1 to simulate the process and retrieve the corresponding output data. For the sake of simplicity, the solid phase concentrations of the components were not considered during the investigation. As mentioned previously, the values of the concentrations, purity and yield were taken during CSS operation and solvent utilisation was later calculated during data refining.

The design space was illustrated using the dside package in python, developed by Sachio [8]. The package screens the data to the KPIs constraints, as defined by the user, and iteratively calculates a suitable alpha shape. This can be displayed in either a 2D or 3D space, with a 0% tolerance, ensuring there lies no violating points inside the shape. The design space is indicated in Figure 1 by the black shaded region, a cultivation of all satisfying points (in orange).

3.4. Creation of ANN models

Python was utilised to build the machine learning models, using NumPy and Pandas for data processing and

TensorFlow, Keras and Scikit-learn packages for building, training and validating the ANN. Multiple models were constructed to evaluate how design-space accuracy improves with increasing sample size and to illustrate the potential for neural networks to inform future sampling batches. The ANN would take the 3 input variables (CMOD1, CFEED3, QI1) and produce 2 outputs: purity and yield. Solvent utilisation was not an output of the ANN due to the way it was retrieved. As shown in Equation 3, values for QI1, QI2, CFEED3 and QB are already known, hence, solvent utilisation is only dependent on the yield in the denominator. Therefore, solvent utilisation can be expressed as a function of yield. If the ANN was to calculate solvent utilisation, the model would be penalising the error metrics twice. The solvent utilisation is calculated after, using the yield values from the ANN models.

Table 3. ANN Hyperparameters,

Features	Values
Sample size	100, 200, 300, 400, 500, 600
Activation function	ReLU
Hidden layers	2
Neuron size	64
Optimiser	Adaptive Moment Estimation (Adam)
Epochs	1000

The hyperparameters mentioned in Table 3 were utilised to build ANN models for samples of varying sizes. A simple, linear activation function was used to encapsulate the behaviour of the KPIs in accordance with the inputs.

The sample was split such that 75% was used for training and rest for validation whilst incorporating standardisation for both the inputs and outputs. Standardisation promotes equal contribution of each variable and provides reliable convergence for algorithms that use gradient descent. The spherical nature of the cost function provides a pathway for more direct and efficient steps for finding the optima.

‘Adam’ (adaptive moment estimation), a type of stochastic gradient descent algorithm, was implemented to efficiently find the minimum of the loss function due to its adaption of both momentum and varying learning rate. The momentum term is the incorporation of an exponential weighted moving average of the past gradients which guides the algorithm consistently towards a minimum.

Mean squared error (MSE) was the chosen loss function to evaluate the accuracy of the models and its predictions and calculated using the Equation 4. It calculates the difference between the actual value and the predicted value of the ANN and heavily penalises larger errors. Most importantly, the MSE function is continuous and differentiable ensuring a stable and efficient optimisation process when using ‘Adam’ optimiser. Each sample size model was trained 5 times and their average was taken to stabilise the error metrics and identify any anomalies.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

3.5. ANN Interpolation

To expand the original dataset, 2^{10} input points are generated, using sobol sequencing. These points are then inputted into the ANNs and the outputs of the ANN are added to the dataset. Hence, a new design space, containing both the original HFM data and the ANN predictions, is generated. In theory, this design space should perform better than the original sample, as it is formed using a larger dataset.

3.6. Optimisation and Machine Learning Toolkit (OMLT)

OMLT [9] is a python package that can reformulates a trained neural network into a Mixed-Integer Linear Program (MILP), which is defined by a set of constraints and variables. OMLT can solve inverse problems by finding optimal inputs that minimise user-defined objective (e.g. cost, error) whilst ensuring the ANN's predicted outputs meet the quality constraints.

Upon the selection of a suitable sample size and its corresponding ANN model, it can be used to optimise single-variable input whilst holding certain parameters constant; this allows the identification of a point that would give the highest outputs with some pre-determined inputs. The objective function is designed to simultaneously minimise the difference between the purity and yield of the optimum point and the operator-defined ideal target of 1 for both metrics. The input parameters are constrained such that the investigative variable, θ_3 , is bounded by inequality constraints (determined by its feasible bounds), while θ_1 and θ_2 are held constant at arbitrary points via equality constraints. α and β represent arbitrary values of the input variables, γ is the lower bound and δ is the upper bound of the design variable.

$$\min f(\mathbf{x}) = (1 - \text{purity})^2 + (1 - \text{yield})^2$$

$$\begin{aligned} \text{s.t. } & 0.95 \leq \text{purity} \leq 1 & \theta_1 &= \alpha \\ & 0.85 \leq \text{yield} \leq 1 & \theta_2 &= \beta \\ & 0 \leq \text{solvent utilisation} \leq 8 & \gamma \leq \theta_3 \leq \delta \end{aligned}$$

3.7. Bayesian Neural Networks (BNN)

Bayesian Neural Networks (BNN) have a similar structure to ANNs, with the network being formed of hidden layers containing weights and biases. However, the difference lies in the fact that the weights and biases are represented as probability distributions, rather than being a specific value. These weights and biases change constantly every time an input is evaluated. As a result, slightly different outputs values are achieved for the same input.

By running the same input 50-1000 times, a probability distribution of outputs is generated for one input. This allows for a probabilistic design space where areas in the input space are characterised based on the probability of satisfying the KPIs. This is advantageous as it gives

operators more information about the reliability of areas in the design space.

3.8. BNN Interpolation

The BNN was trained on 400 samples from the HFM. It receives values of the input parameters and predicts the yield and purity (solvent utilisation was left out due to added complexity). Although the training of this BNN is beyond the scope of the present report, it is important to note that the model was trained on the same dataset used for the 400 sample ANN. For this study, output distribution was produced by sampling the same input 1,000 times. A total of 100,000 Sobol sequenced points in the input space were evaluated to generate their corresponding probability distributions. By assuming the outputs to be normally distributed, the probabilistic design spaces can be formed with the mean and standard deviation of each output.

3.9. Assessing Design Space Accuracy

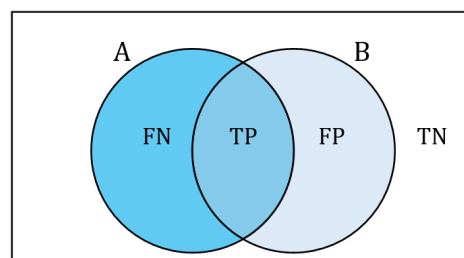


Figure 4. Venn diagram with different error types labelled, where A represents the true design space and B represents the design space of interest

The availability of a HFM allows for a near-true design space to be generated. The accuracy of any design space of interest can then be evaluated by comparing it to this true design space. This comparison must classify the position of both regions. To do so, a dense set of 2^{13} Sobol points is generated and passed through a dside function. The function returns a Boolean value indicating whether each coordinate lies inside or outside the region. By rigorously sampling the input space and identifying which points fall within the boundaries, the position of each design space is captured as a Boolean array. Once these arrays are obtained: one for the true design space and one for the design space of interest, they can be directly compared to quantify how similar the two regions are. Two measures were used to quantify the design spaces: the Jaccard Index and the Accuracy.

The Jaccard index is a measure of similarity between two sets. It is useful as a measure because it captures both false positives and false negatives errors, both of which are detrimental for deterministic design space accuracy. It is intuitively defined and calculated using Equation 5 by classifying coordinates as TP, TN, FP, FN. It ranges from 0-1 with 1 representing equivalence between the two sets and 0 representing no similarity.

In this context false positives are arguably more problematic than false negatives: missing a potentially good

region is less damaging than incorrectly identifying an infeasible point as feasible. Accuracy, defined in Equation 6, directly measures how much of the predicted feasible region lies within the true design space. This is especially important for probabilistic design spaces, which consist of multiple nested regions that need to be assessed differently. While the Jaccard Index remains appropriate for lower confidence bounds, where the aim is to enclose the entire feasible region, it becomes less suitable for higher bounds that are intended to isolate a smaller, more reliable region. Thus, accuracy as a metric is more suitable because it is better aligned with the purpose of higher confidence regions.

$$JI = \frac{A \cap B}{A \cup B} = \frac{TP}{TP + FP + FN} \quad (5)$$

$$\text{Accuracy} = \frac{A \cap B}{B} = \frac{TP}{TP + FP} \quad (6)$$

4. Results and Discussion

4.1. Initial Space

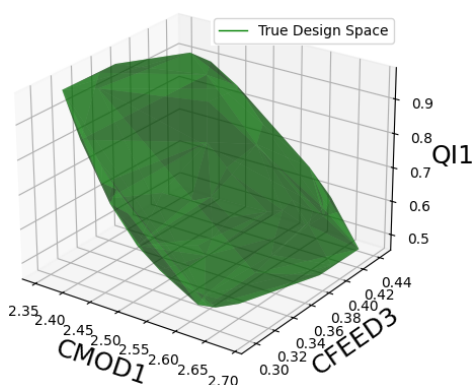


Figure 5. True design space with KPI constraints of 95% purity, 85% yield and 8 mL/mg solvent utilisation.

Understanding of the MCSGP process allows the shape of the resulting design space to be interpreted intuitively. In Figure 6, the feasible values of the system flow rate in phase I1 (QI1) and the solvent strength (CMOD1) appear inversely related. This is consistent with process behaviour: both parameters accelerate the elution of species, so an increase in one can be compensated by a reduction in the other to maintain the desired separation. Consequently, if increasing the value of one causes the process to exit the design space, it can re-enter by reducing the other variable.

Product concentration can also be seen to influence the operating region as demonstrated in Figure 5. Higher concentrations tend to provide greater flexibility, leading to an expansion of the feasible region as the product concentration increases.

4.2. Probabilistic Design Space

The 3D probabilistic design space shown in Figure 7 resembles the deterministic design space in terms of the overall shape and the negative correlation between feasible modifier concentration and the Phase I1 flow rate. As the confidence level increases, the feasible region contracts uniformly.

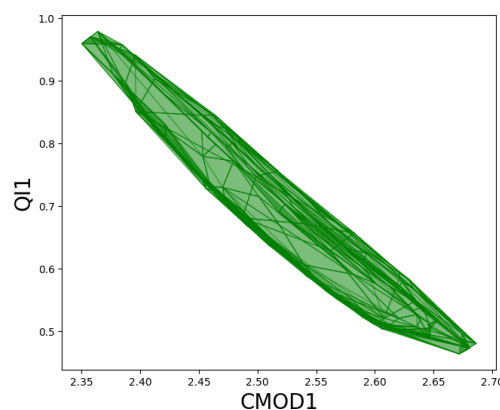


Figure 6. Side profile of true design space when CFEED3 is fixed to show relationship between QI1 and CMOD1.

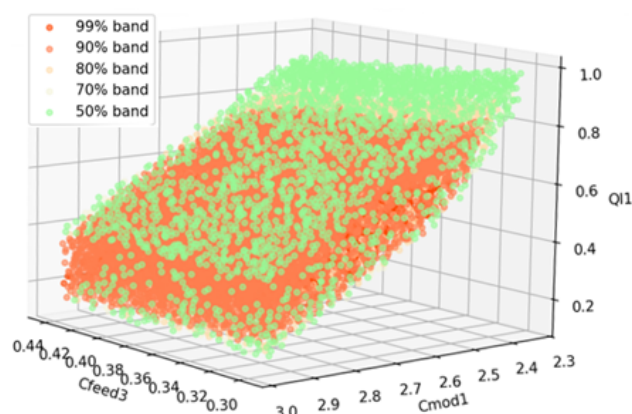


Figure 7. Probabilistic design space with varying confidence bounds from 50% to 90%, each indicated by different colours.

The 3D design space is useful for a broad qualitative understanding. However, it is highly populated and visually dense, making it difficult to identify optimal ranges. This limits its use as a tool for operators and makes it hard to make any judgements on the design space accuracy.

4.3. Assessing the Performance of ANNs

Table 4. Error metric calculations of different ANN models.

Sample size	MSE	MAE	Purity R ²	Yield R ²
100	0.0188	0.0840	0.8834	0.8517
200	0.0021	0.0266	0.9905	0.9806
300	0.0009	0.0180	0.9938	0.9935
400	0.0007	0.0145	0.9938	0.9963
500	0.0007	0.0124	0.9924	0.9983
600	0.0003	0.0107	0.9979	0.9976

Each ANN was built using sample sizes from 100 to 600, and their accuracy was evaluated using the MSE and MAE values in Table 4. As expected, the 600 sample model achieved the lowest errors, with MAE and MSE reduced by 86% and 97% compared to the 100 sample model. This reflects the ANN's improved ability to capture the underlying process behaviour as more data become available, while diminishing returns suggest that additional samples (e.g., 700 or 800) would produce only marginal

improvements.

The coefficients of determination (R^2), also shown in Table 4, further confirm this trend. Purity consistently shows higher R^2 values than yield, and after 300 samples both KPIs exhibit R^2 values above 0.99, indicating that the ANN accurately reproduces the behaviour of the MCSGP process.

Overall, these metrics demonstrate that the chosen architecture, two hidden layers with ReLU activation and 64 neurons, was sufficient to learn the input-output relationships of the system. This is reasonable given that the model predicts only two KPI values, rather than full elution profiles as in previous work by Michalopoulou[7].

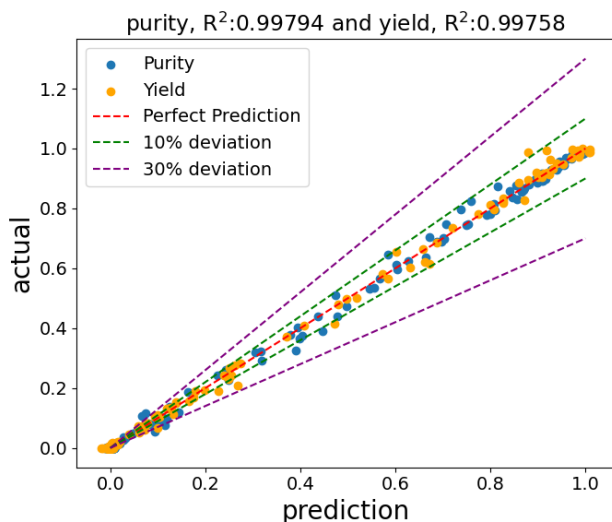


Figure 8. Parity plot for 600 sample trained ANN. It compares the actual and predictive values of yield (orange) and purity (blue). The red lines indicate ideal predictions with 10% (in green) and 30% (in purple) deviation lines.

4.4. Deterministic Design Space Performance

The performance of the ANN models shows they are directly dependent on the sample size. As the training dataset increases from 100 to 600 samples, both the Jacard Index (JI) and the accuracy show stable growth, indicating a better representation of the true design space progressively. The most substantial increase happens between 100 and 500 samples, after which the improvement becomes marginal and the JI increases by only 3.02% from 500 to 600 samples (Table 5). The best performing ANN model was the one trained on 600 samples which

achieved a JI of 0.65566 and an accuracy of 79.75%.

On the other hand, the sample-based design spaces behave differently. Their accuracies remain close to 1 across all sample sizes, as they are generated deterministically and involve no model uncertainty. However, the corresponding JI values are consistently much lower than those of the ANNs. Even with 600 samples, the sample design space captures only 54.99% of the true region. This behaviour reflects the inherent sparsity of the sample method: the sample design space highlights regions that are certainly feasible but fails to capture large areas of the true space. In this sense, the sample-derived design spaces resemble the high-confidence layers of a probabilistic design space highly precise but extremely conservative.

The ANNs consistently outperform their own respective sample design spaces. On average, ANN JIs were 26.89% higher than their corresponding sample JIs, which demonstrated the model's ability to interpolate between sparse points and explore the design space more effectively. The most pronounced difference was at 200 samples, where the ANN provides a 17.7 times greater design space representation than the one produced using the sample. This highlights how ANNs can leverage limited experimental data to infer the underlying kinetics and non-linearity of the MCSGP process and create the design space boundaries effectively.

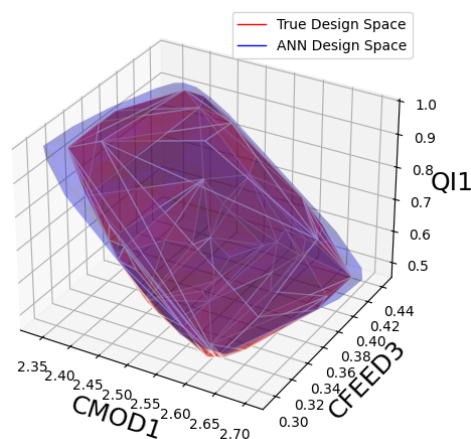


Figure 9. Comparison and overlap of the design space defined by ANN (in blue) and the true design space (in red).

Figure 9 further illustrates how the 600-sample ANN encapsulates the true design space. However, the model slightly overestimates in certain boundary regions, lead-

Table 5. Comparison of ANN and sample performance across different sample sizes.

Sample Size	ANN				Sample			
	JI	Accuracy	FP/FN	DS Size [%]	JI	Accuracy	FP/FN	DS Size [%]
100	0.327814	0.508658	1.117478	29.87	0.035294	1	0	1.49
200	0.457644	0.619971	1.132074	35.26	0.025882	1	0	32.16
300	0.531495	0.587955	4.183936	57.68	0.145882	1	0	33.00
400	0.571534	0.681977	2.552381	74.10	0.387138	0.982821	0.010837	44.69
500	0.625779	0.682418	3.882113	102.98	0.412333	0.984416	0.011296	46.29
600	0.655657	0.796583	1.255283	89.65	0.549877	0.984237	0.020405	59.17

ing to false positives. These arise because small prediction errors near KPI cut-off values cause disproportionately large shifts in the inferred design-space boundary, reflecting the model's sensitivity in these regions. This behaviour is also seen in the parity plot in Figure 8. Purity predictions remain within a 10% deviation band, but yield predictions particularly at low values show deviations exceeding 30%, caused by the sharp KPI fluctuations outside the feasible region. This combination of high accuracy but lower JI reflects the boundary overestimation observed in the ANN model. As discussed later (Section 4.8), this issue is mitigated when using a classification-based ANN rather than a regression model.

As previously discussed, false positives are more problematic than false negatives, since incorrectly identifying infeasible points as feasible carries operational risk. However, this behaviour also suggests a practical role for ANNs: guiding targeted batch sampling. A viable workflow would involve generating an initial experimental batch, training an ANN, constructing the ANN-derived design space, and then sampling the next batch deliberately along the predicted boundaries. This iterative approach refines the design space where uncertainty is highest, enabling more efficient use of experimental resources and improving overall design space accuracy.

4.5. Probabilistic Design Space Performance

Table 6. Performance metrics at different confidence levels and design spaces for 90% purity and 90% yield.

Confidence Level [%]	JI	DS Size [%]	FP/FN	Accuracy
50	0.489	150.9	3.94	0.545
60	0.463	133.5	2.29	0.553
70	0.246	42.56	0.187	0.676
80	0.160	21.73	0.0604	0.770
90	0.00378	0.76	0.000584	0.867
400 Sample ANN	0.748	87.13	3.47	0.783
Sample Design Space	0.661	71.12	0.0802	0.963

The KPI values for the probabilistic design space were reduced to 90% purity and 80% yield. Without this reduction, there would be too few points with 90% confidence of satisfying the KPI requirements to make a successful design space.

As previously mentioned, a probabilistic design space can be viewed as a set of nested regions corresponding to different confidence levels. Since higher confidence bounds aim to highlight the most reliable operating region, the accuracy is a more appropriate metric for evaluating their performance.

The general trend in accuracy is that it increases proportionally with the confidence level. This reflects well on the BNN as it shows its confidence levels to be aligned with the design space accuracy.

When comparing confidence regions to deterministic design spaces, the fairest comparison can be made by looking at the 50% confidence region. This is because the 50% confidence region shows all the outputs whose mean values are above the KPI requirements. It does not account for the standard deviation of output distributions,

making it equivalent to taking the mean value of an ANN that has been sampled and retrained 1000 times.

Unfortunately, the Jaccard Index for the 50% confidence region is worse than that of the original sample and the ANN trained on the same 400 data points. In addition to this, the ANN has a higher accuracy than all but the 90% confidence region, indicating it has outperformed it in both metrics. This can be explained by the exclusion of the original dataset from the probabilistic design space. As a result, the BNN is not interpolating between existing data points, but rather forming a new design space completely from scratch. The differences between the ANN and sample values in Tables 6 and 5 arise from the different KPI values used to generate their respective design spaces.

4.6. 2D Probabilistic Design Spaces

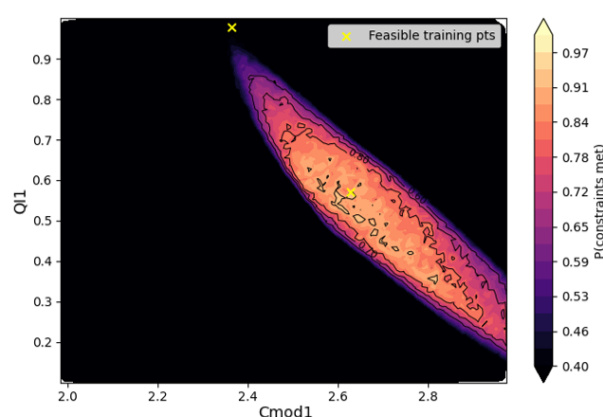


Figure 10. 2D probabilistic design space and CFEED3: 0.31, with the feasible sample points marked with a yellow cross.

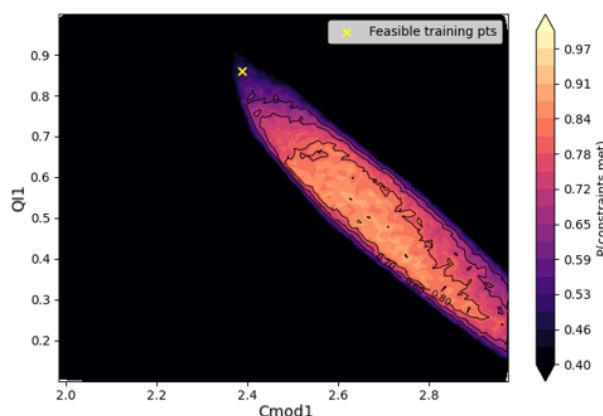


Figure 11. 2D probabilistic design space and CFEED3: 0.38, with the feasible sample point marked with a yellow cross.

Despite the low performance of the probabilistic design space, there is still intrinsic value in the additional information it provides about the input space. The following analysis of 2D design spaces demonstrates this point.

As an operator selecting input conditions for a process at a given product concentration, it is better to choose

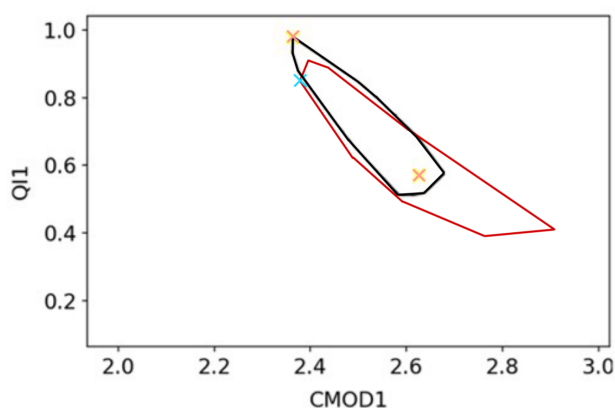


Figure 12. 2D true design space at CFEED3: 0.38 (red) and CFEED3: 0.31 (black) with feasible operating points at each concentration marked in blue and yellow respectively.

an operating point drawn directly from the experimental dataset rather than an untested point that simply falls within the highest-confidence region. Experimental points can effectively be treated as 100% certain, whereas BNN predictions always carry some uncertainty.

Nonetheless, a probabilistic design space still offers valuable insight. It allows operators to assess the flexibility of the experimental points and can guide decisions about where additional experiments might be most informative. In Figure 10, for example, two experimental points exhibit similar KPI values, yet the probabilistic design space reveals that one lies within a higher-confidence region. This distinction provides a clearer basis for selecting the more reliable operating point.

In Figure 11, the only experimental point lies in a relatively low-confidence region. This signals to operators that further exploration of the surrounding high-confidence areas may be worthwhile, both to verify feasibility and to potentially identify more reliable operating conditions. When compared to the true design space (shown in Figure 12), the predictions regarding the position of the operating points in the design space are seen to be true. This shows that even for a low-accuracy probabilistic design space, there is still meaningful value.

These cases show that a great advantage of the probabilistic design space lies in the visual understanding it provides. This is important to consider when comparing BNNs and ANNs, because the additional insight offered by a probabilistic design space cannot be captured in a single metric.

4.7. Implementing ANNs to OMLT

Accuracy of the OMLT formulation was assessed to justify its viability as an alternative for operators. Arbitrary values were selected for CFEED3 and QI1 and the design variable, CMOD1, was implemented into OMLT, for which it gave a prediction of 2.49781. These values were then tested using the deterministic model and its outputs were recorded in Table 7. The true value of yield was close to the OMLT prediction with only 2.04% deviation. However, the repercussion was yield having a 13.75% deviation, which might indicate a bias of OMLT towards

purity instead.

This error can be attributed to the inherent ANN model itself as it had a lower coefficient of determination for yield than purity, meaning its estimations of purity is more accurate which is reflected in the OMLT prediction. As mentioned previously, solvent utilisation is expressed as a function of yield which explains the similar deviation for both KPIs. However, the outputs from the OMLT still satisfy the KPI constraints, hence it is considered as an acceptable prediction. One way to overcome this is via increasing the weighted sum term on yield in the objective function. As of now, the solver places equal importance on both KPIs, but this adjustment would force the solver to recognise and focus on improving the yield error to obtain better predictions.

The specific use for operators of this OMLT formulation is to select points which they have already experimentally tested. As the results of this trial is to be 100% certain, they can optimise these input variables further. Using the knowledge learnt from existing experimental trials, OMLT can provide guidance into the right directions for operators to look at. If a selected point shows promise towards the desired constraints, they can individually optimise each variable, encourage the exploration of the design space and specifically identify regions with better performing KPI values.

Table 7. Comparison of OMLT and actual parameter values.

Parameter	OMLT	Actual	% Error
CMOD1	2.49781	2.49781	-
CFEED3	0.43022	0.43022	-
QI1	0.66056	0.66056	-
Purity	0.99993	0.979543	2.04
Yield	0.99993	0.862453	13.75
Solvent Utilisation	2.00042	2.319291	15.94

4.8. Neural Network Accuracy vs. Design Space Accuracy

It is essential to recognise that regression model accuracy is very different from design space accuracy. All ANN models achieved reasonable error metrics during validation and displayed excellent coefficients of determination for each KPI, showing that the surrogate models captured the underlying process behaviour well. However, this does not translate into accurate design space performance. For example, the 200-sample ANN model, despite having coefficients of determination close to 1, achieved only a JI of 0.4576 and an accuracy of 62%. For good design space performance, the model only needs to predict KPI values accurately at their respective cut-off points. Small errors near these cut-offs therefore matter far more than errors elsewhere, meaning that regression metrics fail to reflect the errors that actually affect design space classification. For this reason, classification models are arguably better aligned with design space accuracy.

5. Conclusion

This paper has demonstrated how neural networks can be used to aid in design space generation. With ANNs consistently outperforming the samples with a greater average Jaccard Index of 0.267 and BNNs showing great promise in enhancing process understanding through probabilistic design spaces. The findings of the report show that there is great potential for NNs to be applied to batch sampling and optimisation problems.

The simple ANN parameters were able to accurately capture the relationship between the decision variables and the KPIs, where a high accuracy of 79.7% was recorded with 600 samples. Although the JI of BNNs were unsatisfactory, importance should be placed on its peak accuracy of 86.7% at 90% confidence bound where operators can identify regions with profound confidence and have a greater chance of satisfying constraints compared to ANNs. Despite this, ANNs were seen to largely outperform BNNs during this study, but this can be attributed to the BNN design space not including the original dataset it was trained on.

Additionally, the study has revealed a clear distinction between model accuracy and design space accuracy (in the context of regression models). All ANNs had reasonable error metrics, yet the design space performance did not fully reflect this. This demonstrates that regression models do not capture design space error very well.

The use of OMLT was influential in the optimisation of input variables via reverse-search. OMLT proved to be useful in guiding operators to identify regions and make well-informed decisions regarding the process.

While the study presents a range of performances across ANN, BNN, and OMLT approaches, an important outcome of this work not contained in the results is the general framework established for assessing design space performance. This approach is not limited to the MCSGP process. Rather it can be applied to different models, datasets, and processes.

6. Outlook

With the implementation of a simple BNN model, work should be done to create a design space conditioned on the fact that the existing data has a 100% probability of being included. The subsequent designs space produced from this should then be analysed using the same framework. This would theoretically greatly influence the low JI of the BNNs and may increase the accuracy as well.

It is recommended that classification models be explored, for both ANNs and BNNs. While regression models provide richer information by giving KPI magnitudes, classification models may yield more accurate design spaces. This is because their evaluation metrics are more closely aligned with design-space performance. Classification models would ideally make a clearer distinction at the critical threshold for the design space regions and may yield greater results.

The use of OMLT was rather preliminary in this study and should not be limited to 3 dimensions. The exciting prospect of OMLT is to leverage its capacity to study the

higher dimensionality of the MCSGP process and implement a variable for process flexibility. The aim would be to maximise distance between the optimal point and the design space boundary. This approach would ensure that any deviation or variability of the input parameters would not violate the KPI constraints.

7. Acknowledgments

The authors thank George Spencer and Foteini Michalopolou for their continued support and guidance throughout this whole research project.

References

- [1] A. Kasemiire, *Design of experiments and design space approaches in the pharmaceutical bioprocess optimization*. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/34147574/>.
- [2] S. Sachio, *A model-based approach towards accelerated process development: A case study on chromatography*. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0263876223005166>.
- [3] R. Chiplunkar, *Probabilistic design space identification for upstream bioprocesses under limited data availability*. [Online]. Available: <https://psecommunity.org/wp-content/plugins/wpor/includes/file/2506/LAPSE-2025.0563-1v1.pdf>.
- [4] C.-T. Yang, *Big data and machine learning driven bioprocessing – recent trends and critical analysis*. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0960852423000512>.
- [5] M. M. M. Krättli F. Steinebach, *Online control of the twin-column countercurrent solvent gradient process for biochromatography*. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0021967313005529?via%3Dihub>.
- [6] M. M. Lars Aumann, *A continuous multicolumn countercurrent solventgradient purification (mcsgp) process*. [Online]. Available: <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/bit.21527>.
- [7] F. Michalopolou, *Late-stage review: Hybrid modelling, optimisation and control of periodic processes*.
- [8] S. Sachio, C. Kontoravdi, and M. M. Papathanasiou, *Design space analysis*. [Online]. Available: <https://github.com/stvsach/dside>.
- [9] F. Ceccon, J. Jalving, J. Haddad, *et al.*, “Omlt: Optimization machine learning toolkit”, *Journal of Machine Learning Research*, vol. 23, no. 349, pp. 1–8, 2022.

Modelling of Integrated Photovoltaic-Photoelectrochemical System for Hydrogen Production

Angel Tang and Venus Ho

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

A photoelectrochemical (PEC) reactor uses a semiconductor photoelectrode to directly convert solar energy into chemical energy for water splitting. It can be integrated with solar cells to provide additional voltage needed for the reaction. This study developed a methodology of modelling 4 1D solar cells connected in series with a PEC reactor using COMSOL Multiphysics to analyse their performance under time-dependent irradiances. The solar cells are simulated using a semiconductor interface to determine their current-voltage characteristics, while the PEC reactor is modelled using a secondary current distribution framework. A parametric sweep of solar irradiance from 6 am to 7 pm enabled the evaluation of the system throughout the day. The results demonstrate that the PEC reactor imposes a low photocurrent, forcing the PV array to function beyond its maximum power point. 20% of the PV power is lost as heat instead of being supplied to the reactor. The integrated system achieved a low average solar-to-hydrogen efficiency of 3.1%. Loss mechanisms within the photoanode, such as mass transport, charge recombination, and optical attenuation are the main contributors. These findings highlight the importance of improving photoanode performance and PV-PEC load matching for future system optimisation.

Introduction

The global energy landscape has undergone several transformations throughout the past centuries. In the 1780s, coal overtook wood to be the largest share in the primary mix for the first time, with technological advances in coal mining [1]. Development in geological theory, drilling and refining technologies led oil and gas to replace coal as the largest energy source in the 1960s [2]. In recent years, environmental concerns associated with the use of fossil fuels have become evident. Extensive use of coal plays a major role in fog in London in the early 20th century and serious air pollution in parts of China now. Since 2000, global energy-related CO₂ emissions have increased by nearly 50% [3]. It recorded a peak of 37.6 billion metric tons in 2024, with half of the increase from the combustion of natural gas [3]. These impacts have pushed more countries to undergo an energy transition to cleaner fuels.

Hydrogen fuel is seen as the energy solution because of its immense energy capacity, accessibility and zero emissions [4]. It is abundant in water and organic compounds. Only heat and water are produced from its combustion. Its chemical energy can be converted to electric energy by combining with oxygen. Notably, the conversion efficiency of up to 60% [5] is higher than other electrical generation alternatives based on fossil fuels, making hydrogen fuel cells more attractive.

The main challenge is to produce hydrogen without emitting carbon dioxide. Steam Methane Reforming (SMR) is the main method to produce hydrogen nowadays. This process uses heat and high pressure to convert natural gas to hydrogen and emits carbon dioxide as a byproduct [6]. According to the IEA global hydrogen review 2025 [7], nearly two-thirds

of the global supply for hydrogen is produced by SMR, and the low-emission production route only takes up less than 1%. Expanding green hydrogen production is therefore critical.

One possible method to produce green hydrogen is by solar energy. This can be done by connecting solar panels to water electrolyzers. It uses PV-generated electricity to power water electrolysis. Another solar-to-hydrogen pathway is to use a photoelectrochemical (PEC) reactor [8]. A semiconductor made from low-cost and earth-abundant materials is immersed in a water-based electrolyte, where solar energy is converted to chemical energy directly for water splitting. The electrolyte cools the illuminated semiconductor [9], keeping the photovoltage high. Additionally, this method removes the need for the electrical current collection network and concomitant current transmission losses.

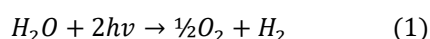
PEC reactors often have limited photovoltage, which is insufficient to drive the water splitting reaction. By connecting PV cells in series, the combined photovoltage output can provide additional driving force for electrolysis.

This study models and simulates a four 1D silicon solar cell module integrated with a 3D PEC reactor using COMSOL Multiphysics. The first objective is to develop and validate the PV model in series configuration and establish its I-V characteristics before coupling with the PEC reactor. The second objective is to analyse the integrated PV-PEC system to determine its operating points throughout the day under realistic variations in solar irradiance. The final aim is to evaluate the performance limitations of the combined system and identify pathways for improvement. While PV and PEC

systems have been extensively studied independently, integrated modelling of a PV array directly coupled to a PEC reactor remains limited in the literature, motivating the approach adopted in this work.

Background

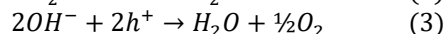
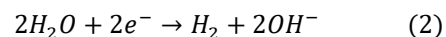
PEC water splitting is a promising solar-to-hydrogen conversion technology where solar energy is directly converted into chemical energy in the form of hydrogen. In a PEC reactor, incident photons generate electron-hole pairs within a semiconductor electrode, and these photogenerated charge carriers drive the water splitting reaction:



A PEC system typically includes an anode and a cathode immersed within an aqueous electrolyte and is often divided by an ion-conductive membrane to prevent the crossover of produced gases [10]. Depending on the design, one or both electrodes serve as photoelectrodes. Although the thermodynamic minimum potential for water splitting is 1.23V at standard conditions, practical PEC operation requires a higher applied potential due to overpotentials associated with charge-transfer kinetics, ohmic resistance, and mass-transport losses.

When a semiconductor photoelectrode comes into contact with an electrolyte, equilibration of their chemical potentials creates band bending and forms a depletion region at the interface. Within this region, a built-in electric field is developed, which is essential for separating the photogenerated carriers. When the semiconductor absorbs photons with energies greater than its band gap, electrons are excited into the conduction band and are driven in opposite directions from the corresponding holes in the valence band, preventing them from recombining [11].

For water splitting to proceed efficiently, the semiconductor band edges must also be properly aligned with the redox potentials of water. The conduction band must be more negative than the proton-reduction potential to enable the hydrogen evolution reaction (HER), whereas the valence band must be more positive than the water-oxidation potential required for the oxygen evolution reaction (OER) [12]. For an n-type photoanode, upward band bending directs holes to the electrolyte to oxidise water, and the electrons are transferred through an external circuit to a p-type photocathode. At the photocathode, downward band bending facilitates electrons to move into the electrolyte, where they reduce protons to form hydrogen [13]. In a basic electrolyte, water splitting involves the following half-reactions:



The concept of using a photoelectrochemical reactor to split water was first demonstrated by Honda and Fujishima in 1972 [14]. A TiO₂ photoanode was used to drive photoelectrochemical water oxidation under ultraviolet illumination to mimic natural photosynthesis. Since then, identifying semiconductor materials capable of visible-light absorption, improving charge separation, and integrating catalytic layers have become a major focus for researchers to improve solar-to-hydrogen efficiency.

Metal oxide semiconductors such as Fe₂O₃, WO₃, TiO₂, ZnO, and BiVO₄ have been studied for photoanode materials for their low cost and earth abundance. In recent years, monoclinic scheelite phase BiVO₄ has emerged as a potential photoanode material. It was first reported to be photoactive for water oxidation in 1998 [15] with a band gap of 2.4 eV, enabling the absorption of visible light, which accounts for 46 % of the sunlight [16]. One drawback exhibited by this material has been the large overpotential required to drive the oxygen evolution reaction (OER).

In addition, most photoanode materials cannot independently achieve water reduction because the photovoltage generated is too low. This can be solved by applying bias voltage provided by economical PV cell devices to improve efficiency. Several demonstrations of fully bias-free water-splitting systems using BiVO₄ photoanodes in tandem with PV cells have been reported. Achieved Solar-to-Hydrogen (STH) efficiencies include 5.2% using double-junction amorphous silicon PV, 5.7% using DSSCs, and 8.1% using GaAs/InGaAsP tandem cells [17], showcasing the potential of PV-PEC hybrids to surpass the voltage limitations of single photoelectrodes.

In the experimental system developed by Dr. Anna Hankin and Dr. Craig McGregor, a photovoltaic-assisted photoelectrochemical (PV-PEC) reactor was constructed to investigate solar-driven water splitting under real outdoor conditions which as shown in **Figure 1**. The setup consists of a BiVO₄ photoanode deposited on fluorine-doped tin oxide (FTO) glass, paired with a titanium cathode, with the two compartments separated by a Nafion membrane to allow ionic transport. A pH 9 borate buffer served as the electrolyte due to its stability and compatibility with BiVO₄. The photoanode was designed in a way that the FTO side faced outward toward the incident sunlight, while the BiVO₄ layer was in contact with the electrolyte. Back-illumination is enabled through the transparent conductive FTO. A set of crystalline silicon (c-Si)

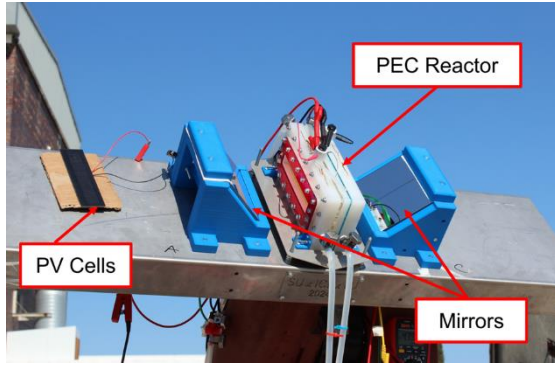


Figure 1. Experimental setup in Stellenbosch University, South Africa

photovoltaic cells was externally electrically connected to the PEC reactor. The entire setting is installed on a dual-axis solar tracking platform to ensure alignment with the sun throughout the day.

This study evaluates the performance of a PV-assisted PEC reactor under time-varying solar irradiance over a full day. This coupled PV-PEC modelling framework advances the understanding of hybrid solar driven hydrogen production and contributes to the development of scalable sustainable energy technology.

Methodology

Semiconductor Module

The silicon solar (PV) cells are modelled using the Semiconductor Physics interface in COMSOL, which solves Poisson's equation for the electric potential and the drift-diffusion equations for electrons and hole transport within the semiconductor.

Poisson's equation governs the electrostatic potential V :

$$\nabla \cdot (-\epsilon_0 \epsilon_r \nabla V) = q(p - n + N_d^+ - N_a^-) \quad (4)$$

where ϵ_0 is the permittivity of free space, ϵ_r is the relative permittivity of the semiconductor, q is the electron charge, n and p are the electron and hole concentrations, N_d^+ and N_a^- are the ionised donor and acceptor concentrations.

Charge conservation for electrons and holes is described by the continuity equations:

$$\frac{\partial n}{\partial t} = \frac{1}{q} (\nabla \cdot \mathbf{J}_n) - U_n \quad (5)$$

$$\frac{\partial p}{\partial t} = \frac{1}{q} (\nabla \cdot \mathbf{J}_p) - U_p \quad (6)$$

where $\mathbf{J}_n$ and $\mathbf{J}_p$ are the electron and hole current densities arising from drift and diffusion. U_n and U_p are the net recombination rates for electron and hole,

as $U_n = U_p = R - G_{ph}$ under steady-state conditions, with R the recombination rate and G_{ph} the photo-generation rate.

The photo-generation rate for electron-hole pairs (G_{ph}) is supplied as a user-defined input variable to the model and is dynamically calculated as a function of depth by performing the spectral integration during the solution process. The photo-generation rate equation is given by:

$$G_{ph}(z) = \int_{250}^{1300} \alpha(\lambda) \phi(\lambda) \exp[-\alpha(\lambda)z] d\lambda \quad (7)$$

where λ is the wavelength, $\alpha(\lambda)$ is the absorption coefficient of silicon obtained from experimental data, z is the depth into the cell from the illuminated surface and $\phi(\lambda)$ is the photon generation rate defined as:

$$\phi(\lambda) = \frac{\lambda}{hc} F(\lambda) \quad (8)$$

here, h is the Planck's constant, c is the speed of light in vacuum, and $F(\lambda)$ is the spectral solar irradiance obtained using the Bird Simple Spectral (SPECTRL2) Model [18] for Stellenbosch, South Africa, on December 21st, 2021, corresponding to the summer solstice. This date is chosen since the site-specific solar irradiance data from the actual experimental period was unavailable. Importantly, this date provides the maximum annual irradiance, enabling the simulation of the solar cells' performance under ideal operating conditions. The solar and silicon absorption spectra are integrated over the wavelength range 250-1300 nm, which corresponds to the region where both the solar irradiance and silicon absorption are significant.

The primary recombination effect is modelled using the Shockley-Read-Hall (SRH) Recombination, which is implemented through the Trap-Assisted Recombination feature. The SRH model accounts for carrier recombination through defect levels within the silicon band gap and assumes that these levels maintain a steady-state occupation. The electron (R_n) and hole (R_p) recombination rates, which must be equal at steady state, are calculated dynamically using the standard SRH expression:

$$R_n = R_p = \frac{np - \gamma_n \gamma_p n_{i,eff}^2}{\tau_p(n + n_1) + \tau_n(p + p_1)} \quad (9)$$

where n and p are the electron and hole concentrations, γ_n and γ_p are the electron and hole degeneracy factors, τ_n and τ_p are the corresponding carrier lifetimes. The effective intrinsic carrier concentration ($n_{i,eff}$) is defined as:

$$n_{i,eff} = \sqrt{N_{c0}N_{v0}} \exp\left(-\frac{E_g - \Delta E_g}{2V_{th}}\right) \quad (10)$$

with N_{c0} and N_{v0} the effective densities of states for the conduction and valence band, E_g the band gap energy and ΔE_g the band gap narrowing voltage. The terms n_1 and p_1 , representing the modified electron and hole concentrations at the defect level, respectively are given by:

$$n_1 = \gamma_n n_{i,eff} \exp\left(\frac{\Delta E_t}{V_{th}}\right) \quad (11)$$

$$p_1 = \gamma_p n_{i,eff} \exp\left(-\frac{\Delta E_t}{V_{th}}\right) \quad (12)$$

where ΔE_t is the energy difference between the defect level and the intrinsic level. The thermal potential is defined as:

$$V_{th} = \frac{k_B T}{q} \quad (13)$$

with k_B the Boltzmann's constant, T the lattice temperature.

To define the silicon p-n junction geometry, COMSOL's doping features are used. Uniform bulk n-doping is introduced across the semiconductor domain using the Analytic Doping Model, with a boundary donor concentration of $N_{D0} = 1 \times 10^{16} \text{ cm}^{-3}$. Front surface p-doping is defined through the Geometric Doping Model, where the Boundary Selection for Doping Profile subnode is employed to specify the surface boundary. This p-region is modelled with a peak acceptor concentration of $N_{A0} = 1 \times 10^{19} \text{ cm}^{-3}$ and uses a Gaussian decay profile into the bulk, corresponding to a junction depth of $d_j = 0.25 \mu\text{m}$. Electrical connections at the front and back surfaces are represented by two Metal Contact features, which specify the boundary conditions for charge transport.

Series Connection of 4 1D Silicon Solar Cells

The four 1D silicon solar cells were connected in series using the Electrical Circuit interface. Each solar cell is modelled as separate component with defined semiconductor physics using a cross-sectional area of 7.5 cm^2 and a depth of $150 \mu\text{m}$. The terminal type of the front metal contact of each cell is set as Circuit (current) and the contact type is set as Ideal Ohmic, meaning it acts as a direct interface between the solar cell and the electrical circuit. The front contact was chosen because it is the region where doping profiles are defined and where voltage develops due to built-in junction potential and photogenerated charge separation. This setting enables the terminal to supply its voltage to an

External I vs. U element and receive the circuit current back into the solar cells.

To connect the four devices in series, they were assigned sequential circuit nodes, where the positive terminal of one device connects to the negative terminal of the next. A Ground node was placed at the start of the circuit loop, and the Voltage Source sweeping from 0 to 3 V using an auxiliary sweep was positioned across the entire chain of solar cells to simulate increasing external bias. This sweep range captures the expected operational voltage window for a four-cell array, as each silicon solar cell typically contributes around 0.6 V under illumination. Additionally, a 10Ω resistor was included in the circuit to simulate resistive losses. This configuration enabled current continuity through all four solar cells while allowing the voltage to build cumulatively across the cells.

A realistic representation of solar cell performance throughout the day is achieved by performing a parametric sweep using spectral irradiance values from the SPECTRL2 model [18] at hourly intervals between 06:00 and 19:00. **Figure 2** presents the corresponding solar spectra used as input for the sweep. This time-resolved irradiance data enables the model to evaluate the dynamic response of the solar cell, and consequently the coupled PEC reactor, under daily variations in solar irradiance.

Coupling the 4-cell PV array with PEC Reactor

The PEC reactor was modelled using the Secondary Current Distribution framework, where charge transport in the solid and electrolyte domains was described by Ohm's law and solved using Laplace's equation for the electrode potential ϕ_s and electrolyte potential ϕ_l . Photocurrent generation at the BiVO_4 photoanode was described by a modified Gärtner–Butler equation incorporating interfacial charge-transfer efficiency and spectral photon absorption. A parametric sweep was applied to the absorbed photon flux $(I_0 - I_x)_\lambda$ to model the effect

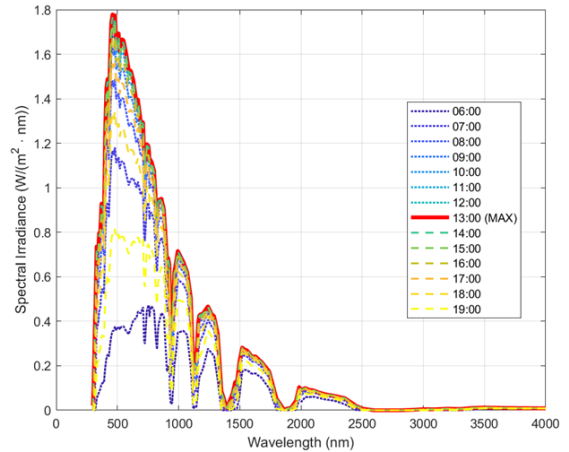


Figure 2. Solar spectra from 06:00 to 19:00 hours in Stellenbosch, South Africa, on December 21, 2021, used as the time-varying input for the parametric sweep.

of hourly solar irradiance on PEC reactor performance. Each parameter value represented the photon flux at a given hour, and because the Gärtner–Butler photocurrent is proportional to $(I_0 - I_x)_\lambda$, the sweep varied the PEC photocurrent supplied to the PV–PEC circuit. This enabled the model to compute the operating point of the coupled system under realistic time-dependent illumination.

The total anodic current I_{photo} is obtained by integrating the local current density over the illuminated area:

$$I_{photo} = \int J_{loc} dA \quad (14)$$

The I–V characteristics of the PV array were imported in the existed PEC reactor model as an interpolation function, with current as the input and voltage as the output. This function was used to define a voltage source in the circuit connecting the PV array to the PEC reactor. The circuit was arranged such that the current generated in the PEC anode matches the PV current, and the corresponding photovoltage was supplied to the PEC system.

I_{photo} was used as the input current to the external PV module through an I–V interpolation function:

$$I_{photo} = I_{PV} \quad (15)$$

$$V_{PV} = f(I_{photo}) \quad (16)$$

The PV voltage output was supplied back to the PEC model through an external circuit node. This ensured that current continuity was preserved while enabling dynamic adjustment of the PV voltage based on the reactor's electrochemical load.

$$V_{PV} = \phi_{s,photoanode} \quad (17)$$

Study Steps and Numerical Solution Strategy

During development, the models frequently encountered convergence errors due to strong non-linear coupling between the four solar cells and the pec reactor. These issues were solved by adopting a structured multi-step solution procedure [19] to ensure numerical stability in the PV array and PV–PEC simulations. The first step was the Semiconductor Equilibrium study, which assumes thermal equilibrium and only Poisson's Equation is solved. This provides built-in potential, depletion region shape, and initial carrier distributions of each solar cell. After equilibrium is obtained, a Stationary study incorporating photogeneration, charge transport, external circuit coupling, and electrochemical boundary conditions was executed.

Because semiconductor and electrochemical kinetics at the boundaries of the PEC reactor model are strongly nonlinear, particularly when multiple PV cells are connected in series. The solver configuration was refined by increasing the maximum number of Newton iterations and enabling damped Newton updates, allowing the solver to proceed with smaller, more stable steps. Intermediate solutions were stored and reused as initial guesses for the dependent variables in the series chain, avoiding convergence failure from starting the entire system simultaneously. Additionally, mesh refinement was applied locally near depletion regions, junctions, and semiconductor–electrolyte interfaces to resolve steep gradients in potential and carrier density.

Assumptions & Model Simplifications

The following assumptions and simplifications were applied to the model:

- The simulation assumes ideal solar tracking, meaning the solar cell is always positioned directly facing the sun to maximise incident irradiance.
- All model parameters are determined at a constant temperature of 293.15K and are detailed in **Table 1**.
- Magnetic fields are not included in the model, simplifying the charge transport
- A local thermodynamic equilibrium is assumed, where the carrier temperature is equal to the lattice temperature, eliminating the need for a separate heat transfer model.
- Complete dopant ionisation is assumed, where $N_d^+ = N_d$ and $N_a^- = N_a$.
- No band gap narrowing is included in the model ($\Delta E_g = 0$).
- The SRH defect energy level is located exactly at the intrinsic level ($\Delta E_t = 0$), which represents the ideal case for maximum recombination activity.

Model Limitations

The PEC reactor model neglects mass-transfer phenomena such as ion diffusion, reactant depletion, pH gradients, convection, and bubble formation, all of which can influence real PEC behaviour and reduce the achievable current at high reaction rates. On the PV side, the connected silicon solar cells are represented through imported I–V curves, an approach adopted for computational simplicity. While this captures the external electrical behaviour of the PV device, it does not resolve internal semiconductor physics such as carrier transport, recombination, spectral sensitivity or thermal effects, within a series-connected PV chain.

Table 1. Parameters used in the COMSOL Semiconductor Module

Symbol	Description	Value	Units
q	Electron charge	1.602×10^{-19}	C
ϵ_0	Permittivity of free space	8.854×10^{-12}	Fm^{-1}
ϵ_r	Relative permittivity of semiconductor	11.7	1
h	Planck's constant	6.626×10^{-34}	$m^2kg s^{-1}$
c	Speed of light in vacuum	2.998×10^8	ms^{-1}
τ_p	Hole degeneracy factors	10	μs
τ_n	Electron degeneracy factors	10	μs
N_{c0}	Effective density of states for the conduction band	2.705×10^{19}	m^{-3}
N_{v0}	Effective density of states for the valence band	1.005×10^{19}	m^{-3}
E_g	Band gap energy	1.12	V
ΔE_g	Band gap narrowing voltage	0	V
k_B	Boltzmann's constant	1.381×10^{-23}	JK^{-1}
T	Temperature	293.15	K
ΔE_t	Energy difference between the defect level and the intrinsic level	0	V

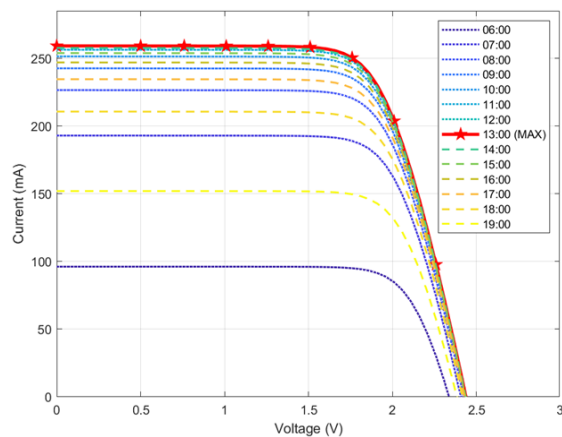
Collectively, these simplifications mean that the PV-PEC model represents an idealised performance, and real devices are expected to exhibit lower efficiencies once transport limitations, full optical losses, and temperature-dependent PV behaviour are considered.

Solar-to-Hydrogen Efficiency

To measure how effectively the PV-PEC system converts solar energy to hydrogen. Solar-to-Hydrogen Efficiency, η_{STH} , is calculated as:

$$\eta_{STH} = \frac{1.23 \times I_{op} \times \eta_F}{P_{solar}} \quad (18)$$

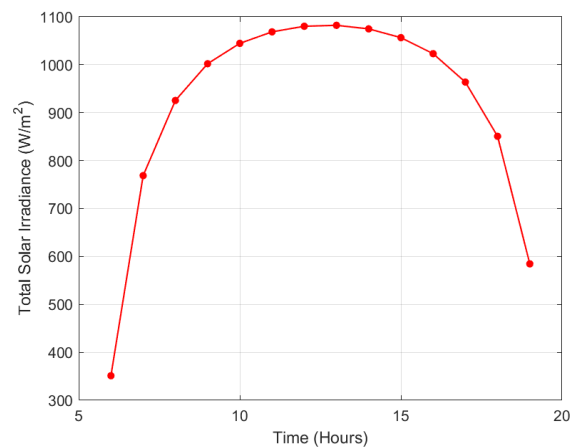
where I_{op} is the operating current, η_F is the Faradic Efficiency, which accounts for charge losses during electrolysis. A value of 84.53% is used based on reported experimental measurements in the literature [20]. P_{solar} is the total solar power input, calculated as the product of the incident irradiance and the total photon-absorbing area (PV area of 30 cm² and photoanode area of 30 cm²).

**Figure 3.** I-V curves of the solar cells connected in series throughout the day between 06:00 to 19:00 hours.

Results and Discussion

PV Array I-V and P-V characteristics

The simulated performance of the four solar cells connected in series, as presented in **Figure 3**, exhibits a family of characteristic I-V curves that correlate with the time-dependent solar irradiance shown in **Figure 4**. The total Solar Irradiance in **Figure 4** is obtained by integrating the spectral irradiance from **Figure 2** across all relevant wavelengths. The pattern observed in the I-V curves mirrors the irradiance time profile, where the curves systematically shift upwards from 06:00, peaking at 13:00, and subsequently drops until 19:00. The 13:00 curve represents the maximum output, starting with a short-circuit current I_{SC} of 259.2 mA. This current value correlates with the simulated peak total solar irradiance of 1082 W/m² from **Figure 4** as the solar cell's current output is almost directly proportional to the incident light intensity. Conversely, the 06:00 curve exhibits the minimum I_{SC} of 96.2 mA, corresponding to the lowest irradiance input of 351 W/m². The open-circuit voltage V_{OC} shows relatively little variation across

**Figure 4.** The total solar irradiance in Stellenbosch, South Africa, on December 21st, 2021.

the day, ranging from ~ 2.33 V to ~ 2.45 V, as V_{OC} has only a logarithmic dependence on irradiance [21].

All I-V curves share the same fundamental shape expected from a silicon p-n junction device. At low voltages, the current remains nearly constant in the short-circuit region because the built-in electric field within the depletion region is sufficiently strong to separate and collect almost all photogenerated electron-hole pairs. The net photovoltaic current I_{PV} , is given by the difference between the photocurrent generated by light absorption I_{ph} and the diode current I_d . Under low bias, I_{ph} dominates and is effectively constant, while I_d is still small, resulting in a flat current plateau. As the applied voltage increases beyond approximately 1.6 V, the diode current rises rapidly, reducing the net current. In this regime, the external bias begins to counteract the built-in electric field, diminishing carrier collection efficiency and increasing recombination losses. Eventually, when I_d equals I_{ph} , the net current I_{PV} becomes zero, corresponding to the open-circuit condition [22].

Figure 5 shows the simulated power-voltage (P-V) curves of the photovoltaic module at hourly intervals from 06:00 to 19:00. The curves illustrate the strong dependence of PV performance on solar irradiance. As irradiance increases toward midday, the maximum power point (MPP) rise. The curve at 13:00, marked with red stars, represents the maximum irradiance condition and therefore exhibits the highest power output across the simulated period. For optimal system performance, the PV module should operate as close as possible to its MPP at any given time. The operating point of the PV-PEC system is fixed by the requirement of equal current through both components. As a result, the PV is forced to operate at the voltage dictated by the PEC reactor. The loss of power dissipates as excess

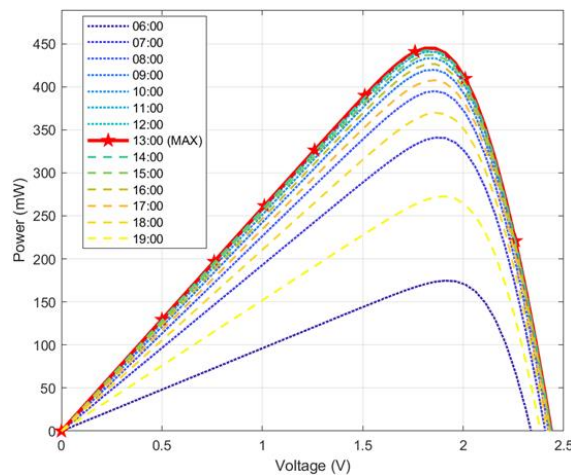


Figure 5. The P-V curves of the solar cells connected in series throughout the day between 06:00 to 19:00 hours.

heat, which needs to be removed to maintain the performance of the cell [23].

Coupled PV-PEC System Performance

Figure 6 illustrates the coupled system solution by plotting the I-V curves of the PV and the PEC reactor respectively at two key time points: minimum irradiance at 06:00 and maximum irradiance at 13:00. The intersection point of the respective solar cells and PEC reactor curves represents the stable electrical operating voltage V_{op} and current I_{op} of the coupled system, which satisfies Kirchhoff's conservation laws and represents the solution point solved for by the COMSOL simulation.

A critical observation from **Figure 6** is the comparison between the system's operating point and the PV's MPP. At minimum irradiance at 6:00, the system operates at $V_{op} = 1.65$ V and $I_{op} = 95.7$ mA corresponding to a power of 157.9 mW, while the MPP is 174.7 mW at $V_{MPP} \approx 1.91$ V. Although the reactor is operated successfully, the voltage remains to the left of the MPP, confirming that the PV array is not utilised at maximum power delivery. Under peak irradiance at 13:00, the PV I-V curve is significantly shifted upward forcing the system to operate at $V_{op} = 2.11$ V and $I_{op} = 168.2$ mA, resulting in an operating power of $P_{op} = 354.9$ mW. For comparison, at this same irradiance the PV's MPP occurs at $V_{MPP} \approx 1.81$ V, delivering 444.7 mW. Thus, despite peak spectral conditions, 20% of available PV power cannot be harvested due to voltage overshoot. With the exception of 06:00, the operating point lies beyond the PV's MPP, forcing the PV system to operate at voltages greater than the V_{MPP} . In this high-voltage, recombination-dominated region, diode current rises sharply and carrier collection becomes inefficient, resulting in lower net operating current and power.

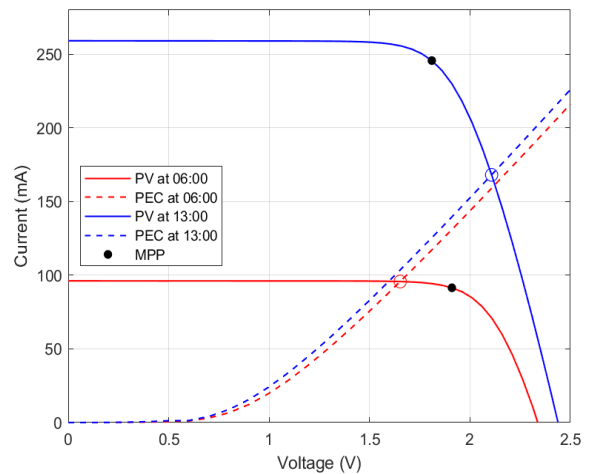


Figure 6. The I-V Curves of PV and PEC Reactor at 06:00 and 13:00 respectively. The operating points are circled and the MPPs are represented by the black circle.

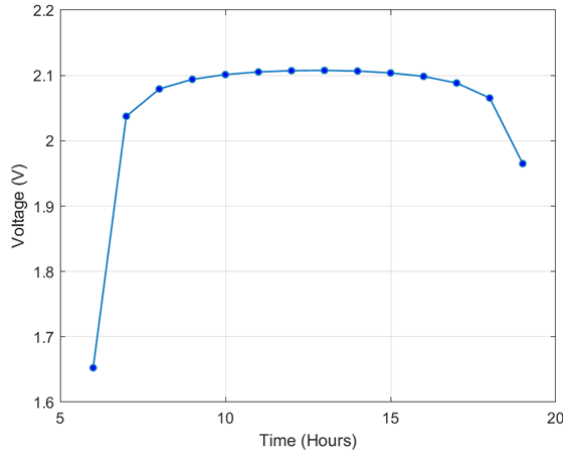


Figure 7. Operating voltage of the coupled PV-PEC system throughout the day.

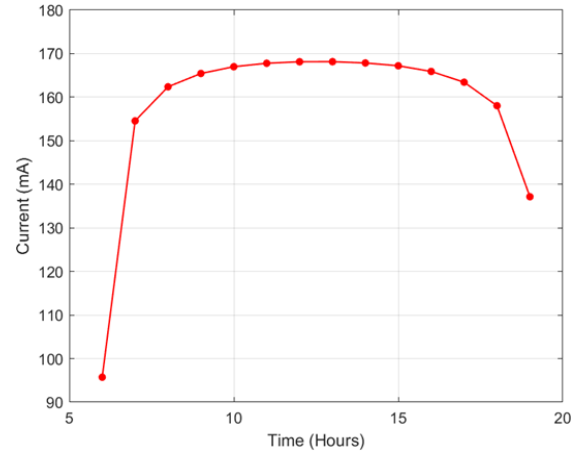


Figure 8. Operating current of the coupled PV-PEC system throughout the day.

The daily variation of the operating voltage and current for the coupled system is presented in **Figures 7 and 8**. Both quantities follow the irradiance trend, with the operating voltage and current increasing from sunrise to a maximum at 13:00, coinciding with peak spectral irradiance, and then declines toward sunset. This behaviour is primarily governed by the PV's response to changing irradiance rather than the PEC reactor. **Figure 6** shows that the PEC reactor's I-V curve changes only slightly between 06:00 and 13:00, indicating that the reactor's behaviour is relatively insensitive to irradiance. In contrast, the PV curve shifts significantly with changing irradiance, causing the operating point to follow the PV's response throughout the day.

The limited variation in the PEC curve results from its rate-limiting electrochemical processes, including slow oxygen evolution kinetics and mass-transport constraints. Therefore, even if solar concentration were applied to the PV to further raise its photocurrent plateau, the operating point would continue to occur beyond the PV's MPP, causing the system to operate inefficiently below its maximum power and providing negligible improvement to overall the STH efficiency. These results indicate that system limitations arise primarily from the PEC reactor rather than insufficient photovoltaic power. Kinetic losses at the oxygen-evolving photoanode and transport limitations within the electrolyte are restricting the performance of the PEC reactor, suggesting future optimisation should prioritise reducing PEC overpotentials, enhancing catalytic performance, and improving mass transport. Based on the operating current throughout the day, the average η_{STH} is approximately 3.1%, indicating that the system remains limited in efficiency even under idealised irradiance and PV conditions.

Comparison with Experimental Data

To validate the model, the simulation outcomes are compared with the same non-concentrated PV-assisted PEC experiment that was reported. Under those conditions, the photoanode received approximately 770 W/m^2 , producing a cell voltage of 2.3 V and a spontaneous photocurrent of 49 mA respectively, which corresponded to a current density of 1.63 mA/cm^2 and an overall η_{STH} of 1.2% [19]. The experimental irradiance closely matches the simulated 07:00 case, where total irradiance is 768 W/m^2 . The operating point for this condition is shown in **Figure 9**. The model predicts that $V_{op} = 2.04 \text{ V}$ and $I_{op} = 154.6 \text{ mA}$, giving a current density of 5.15 mA/cm^2 and a η_{STH} of 3.5%. Importantly, the experimental point aligns with the PV's I-V characteristic at 2.3 V, but it does not lie near the simulated operating point.

The significantly lower photocurrent obtained in the experimental PEC system can be primarily attributed to practical limitations of the BiVO_4 photoanode. BiVO_4 has a short carrier length

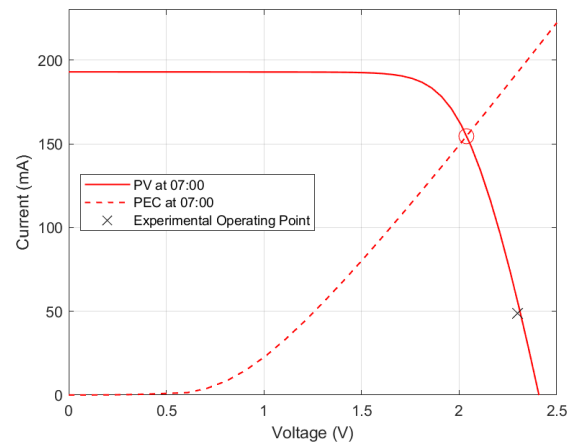


Figure 9. The I-V curves of PV and PEC Reactor at 07:00 with the experimental operating point at similar irradiance shown as the black cross.

of 10 nm and a low carrier mobility of $4.4 \times 10^{-2} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ [24], causing poor bulk charge separation. Photogenerated electron-hole pairs recombine rapidly before they reach the semiconductor-electrolyte interface. In addition, inefficient surface charge injection into the electrolyte limits the rate of the oxygen evolution reaction, causing hole accumulation and further enhanced surface recombination. Moreover, photoanode stability is affected by alkaline electrolyte during the PEC water splitting. In the experiment, where excessive bias is supplied by the PV array, high anodic potential further accelerates oxidation of Bi^3 even with a co-catalyst NiFeOOH to provide more active sites for holes to drive the reaction [20]. This reduced the effective active area and led to a progressive decline in photocurrent during the operation.

The achievable photocurrent was also suppressed by weather conditions and optical losses. Unlike the idealised simulation conditions, the experiment is subject to fluctuating solar irradiance driven by atmospheric conditions and transient cloud cover, lowering the photon flux available. Additionally, photon absorption on the photoanode was reduced due to reflection at the electrolyte-bubble interface and scattering from surface roughness. The presence of bubbles on the photoelectrode lowers the area-averaged light absorption by up to 18% compared to an uncovered photoelectrode [25]. Consequently, the number of photons absorbed by the photoanode and the rate of photogenerated charge carriers available for water oxidation were greatly reduced.

In contrast, the PEC model does not fully capture these loss mechanisms. Mass transport limitations and bubble coverage are neglected in the stationary simulation. Furthermore, optical losses are represented by a constant 10% attenuation of the incident photon flux. This underestimates the cumulative and dynamic losses present in the experimental system. As a result, the model represents an optimistic upper performance bound, leading to an overprediction of the photocurrent relative to experimental observations.

Conclusions and Future Outlook

This study developed and analysed a coupled PV-PEC system by modelling four silicon solar cells connected in series and integrating them with an existing 3D PEC reactor model. The work established the I-V characteristics of the PV array, evaluated the operating points of the coupled system across realistic daily irradiance conditions, and identified the primary factors limiting performance. The results reveals a persistent voltage mismatch, with the PEC reactor forcing the PV to operate beyond its maximum power point for most of the

day. This led to reduced power transfer and a low overall solar-to-hydrogen efficiency with an average value of approximately 3.1%. Importantly, the analysis shows that performance limitations arise mainly from the PEC reactor rather than insufficient photovoltaic power.

Although the existing PEC model provides valuable insight, it can be made more predictive by incorporating more realistic kinetic and transport losses, including bubble-induced mass-transport resistance, electrolyte depletion, and detailed overpotential contributions. At the device level, improvements to the photoanode such as using more active oxygen-evolution catalysts, selecting materials with lower kinetic barriers, and employing electrolytes with reduced resistance and higher ionic conductivity would help increase the achievable current and move the system toward a more efficient operating region. Future work should also investigate PV sizing and load-matching strategies to reduce voltage mismatch and enable more effective utilisation of available solar power.

Overall, the modelling framework developed in this study provides a foundation for future optimisation studies and offers a cost-effective approach for exploring design improvements in coupled PV-PEC systems.

Acknowledgments

The authors would like to thank Sai Muthumaran for their guidance and technical support throughout the modelling process.

References

1. Zhao, Q. *et al.* (2014) 'Energy revolution: From a fossil energy era to a new energy era,' *Natural Gas Industry B*, pp. 1–11
2. Melsted, O. and Pallua, I. (2018) 'The Historical Transition from Coal to Hydrocarbons: Previous Explanations and the Need for an Integrative 2 Perspective,' *Journal of History*, 53(3), pp. 395–422
3. Statista (2025) *Global CO₂ emissions related to energy 1975-2024*. <https://www.statista.com/statistics/526002/energy-related-carbon-dioxide-emissions-worldwide/?srsltid=AfmBOopUATfwB0qBCUS0rWgaDwLk-5QDdJWcRhu8Pt4dDCfh59cptyf7>.
4. Crabtree, G.W. and Dresselhaus (2008) 'The hydrogen fuel alternative,' *MRS Bulletin*, 33(4), pp. 421–428. <https://doi.org/10.1557/mrs2008.84>.

5. Fuel cells (no date). <https://www.energy.gov/eere/fuelcells/fuel-cells?>
6. Hydrogen production: Natural gas reforming (no date). <https://www.energy.gov/eere/fuelcells/hydrogen-production-natural-gas-reforming>.
7. International Energy Agency (2025) *Global Hydrogen Review 2025*.
8. Moss, B. *et al.* (2021) 'A review of inorganic photoelectrode developments and reactor Scale-Up challenges for solar hydrogen production,' *Advanced Energy Materials*, 11(13). <https://doi.org/10.1002/aenm.202003286>.
9. Johnson, E.F. and Haussener, S. (2024) 'Thermal-integration in photoelectrochemistry for fuel and heat co-generation,' *Sustainable Energy & Fuels*, 8(18), pp. 4199–4212. <https://doi.org/10.1039/d4se00304g>.
10. Minggu, L.J., Daud, W.R.W. and Kassim, M.B. (2010) 'An overview of photocells and photoreactors for photoelectrochemical water splitting,' *International Journal of Hydrogen Energy*, 35(11), pp. 5233–5244. <https://doi.org/10.1016/j.ijhydene.2010.02.133>.
11. Butler, M.A. and Ginley, D.S. (1980) 'Principles of photoelectrochemical, solar energy conversion,' *Journal of Materials Science*, 15(1), pp. 1–19. <https://doi.org/10.1007/bf00552421>.
12. Kumar, M. *et al.* (2022) 'Recent trends in photoelectrochemical water splitting: the role of cocatalysts,' *NPG Asia Materials*, 14(1). <https://doi.org/10.1038/s41427-022-00436-x>.
13. Ouardi, M.E. *et al.* (2023) 'Review of photoelectrochemical water splitting: From quantitative approaches to effect of sacrificial agents, oxygen vacancies, thermal and magnetic field on (photo)electrolysis,' *International Journal of Hydrogen Energy*, 51, pp. 1044–1067.
14. Fujishima, A. and Honda, K. (1972) 'Electrochemical photolysis of water at a semiconductor electrode,' *Nature*, 238(5358), pp. 37–38. <https://doi.org/10.1038/238037a0>.
15. Simelys Hernández, Kristine Rodulfo Tolod and Nunzio Russo (2017) 'Recent Advances in the BiVO₄ Photocatalyst for Sun-Driven Water Oxidation: Top-Performing Photoanodes and Scale-Up Challenges,' *Catalysts* [Preprint].
16. Li, F., Li, Z. and Shang, X. (2024) 'Recent surficial modification strategies on BiVO₄ based photoanodes for photoelectrochemical water splitting enhancement,' *International Journal of Hydrogen Energy*, pp. 611–624.
17. Sivula, K. and Van De Krol, R. (2016) 'Semiconducting materials for photoelectrochemical energy conversion,' *Nature Reviews Materials*, 1(2). <https://doi.org/10.1038/natrevmats.2015.10>.
18. Holmgren, W.F., Hansen, C.W. and Mikofski, M.A. (2018) 'pvlib python: a python package for modeling solar energy systems,' *The Journal of Open Source Software*, 3(29), p. 884. <https://doi.org/10.21105/joss.00884>.
19. *Improving convergence of semiconductor models* (no date). <https://www.comsol.com/support/learning-center/article/69861>.
20. Hankin, A. *et al.* (2025) 'Integrating concentrated solar power with a prototype photoelectrochemical reactor,' *Nature Portfolio* [Preprint]. <https://doi.org/10.21203/rs.3.rs-7963830/v1>.
21. *Effect of light Intensity | PVEDucation* (no date). <https://www.pveducation.org/pvcdrom/solar-cell-operation/effect-of-light-intensity> [Accessed 10 Dec. 2025].
22. Atonometrics (2025) *What is a PV Module IV Curve? - Atonometrics*. <https://www.atonometrics.com/application/s/what-is-a-pv-module-iv-curve/> [Accessed 5 Dec. 2025].
23. Rothschild, A. and Dotan, H. (2016) 'Beating the Efficiency of Photovoltaics-Powered Electrolysis with Tandem Cell Photoelectrolysis,' *ACS Energy Letters*, 2(1), pp. 45–51. <https://doi.org/10.1021/acsenergylett.6b00610>.
24. Park, E., Yoo, J. and Lee, K. (2025) 'Ideal structures of isolated BiVO₄ nanoparticles on WO₃ nanocorals to enhance photoelectrochemical hydrogen evolution,' *International Journal of Hydrogen Energy*, pp. 559–568.
25. Bhanawat, A., Zhu, K. and Pilon, L. (2022) 'How do bubbles affect light absorption in photoelectrodes for solar water splitting?,' *Sustainable Energy & Fuels*, 6(3), pp. 910–924. <https://doi.org/10.1039/d1se01730f>.

Biodegradable Microcapsule Formation for Analgesic Delivery via Acoustic Levitation

Anam Ahmed and Nazifa Tabassum

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

For effective acute pain relief, rapid dissolution of paracetamol is essential but its inherent crystalline nature limits dissolution rates. Therefore, encapsulation within an amorphous excipient can accelerate release. Conventional excipients are often petrochemical based, non-biodegradable and costly. Therefore, this study aims to investigate the development of an encapsulated microcapsule using biodegradable and renewable alternatives and study the effects of polymer concentration, drying kinetics, Active Pharmaceutical Ingredient (API) loading on the resulting structure of the encapsulated particle and its dissolution in water and simulated GI fluids. Hydroxypropyl methylcellulose (HPMC) and Cellulose Nano Crystals (CNC) were investigated as candidate excipients. SEM and XRD analysis showed HPMC produced amorphous particles, while CNC demonstrated crystalline behaviour. HPMC-paracetamol solutions were initially prepared in water as and then diluted with various organic solvents. Droplets from these solutions were then dried using the Acoustic Levitator and monitored via optical imaging. Drying curves were generated and revealed that the use of organic solvents significantly reduced the drying time needed. The drying curves were then fitted to an empirical equation to extract key insights on the drying kinetics. Dried particle structures were then analysed using SEM, FTIR and XRD. It was observed that while pure paracetamol exhibited crystalline behaviour, encapsulating it with HPMC formed predominantly amorphous particles with varying residual local ordering. The effect of the key parameters from the empirical equation governing drying behaviour on the XRD peak intensity was then analysed and it was observed that slower drying of the levitated particles and greater shrinkage of them promoted crystallinity. Dissolution tests in water and simulated gastrointestinal (GI) fluids demonstrated that amorphous particles, especially those dried using acetone, exhibited the fastest and most complete dissolution profiles.

Introduction

Rapid pain relief is crucial upon ingestion of analgesic medications. Standard immediate-release tablets typically take 20–30 minutes to alleviate pain, as the drug must first dissolve after tablet disintegration. This delayed onset of action may be fatal in acute pain situations and could potentially lead patients to excessive dosing if they do not experience relief within a short timeframe[1]. By improving the dissolution rate of such drugs, the time to analgesic effect can be shortened and therapeutic outcomes enhanced. Modern drug delivery strategies have increasingly emphasised excipient design to address such challenges. API bioavailability and overall efficacy can be improved by enhancing excipient-controlled properties, such as drug solubility, stability, and release kinetics [2]. By varying excipient formulations, it is possible to design drug-loaded microspheres that release the active ingredient more rapidly and consistently, thereby addressing the issues of delayed onset and variable dosing effectiveness in current painkillers. One approach to achieve faster release is to create microscale drug capsules with tailored excipient matrices. Doing so allows control over the drug's physical form and dissolution profile in the gastrointestinal tract. Conventional industrial production typically involves spray drying; however, this consists of multiple droplets and wall interactions in the drying chamber, making it difficult to monitor individual particle formation. In

this project, we employ acoustic levitation as an *in situ* single droplet drying technique, which serves as a controlled analogue of spray drying. In acoustic levitation, a single droplet is suspended in a node created by standing sound waves, eliminating container contact. This setup enables precise control of drying conditions and real-time observation of individual crystal behavior and formation. Notably, acoustic levitation has been used in pharmaceutical research to rapidly screen amorphous solid dispersions on the single-droplet scale [3]. By levitating droplets of drug-polymer solutions, we can directly observe drying kinetics, particle morphology development, and crystallinity changes without the confounding effects present in bulk spray dryers. This method provides insight into how formulation variables impact the final particle, which can later be scaled up to commercial spray drying processes. The overarching aim of this paper is to leverage this single-particle approach to develop an improved oral drug delivery system that achieves immediate drug release and faster onset of action, while using safer biodegradable excipients.

Background

Hydroxypropyl methylcellulose (HPMC) is a hydrophilic polymer widely employed in oral formulations due to its versatility as a binder, film-former, and matrix agent. In immediate-release tablets, low-viscosity grades of HPMC hydrate

quickly, allowing rapid drug release without retarding dissolution. Hence, HPMC is able to accommodate high drug loading and is compatible with APIs like acetaminophen (paracetamol). Beyond its role in capsule formation, HPMC is recognised for inhibiting drug crystallisation and stabilising amorphous solid dispersions, thereby enhancing the apparent solubility of poorly water-soluble drugs. By restricting molecular mobility and interacting with drug molecules, it maintains supersaturation during dissolution, hence improving dissolution rates and bioavailability. HPMC is also non-toxic and, as a cellulose derivative, is considered biodegradable and safe for ingestion. This makes HPMC an effective excipient for fast-release oral drug delivery systems.

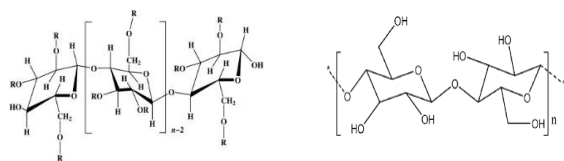


Figure 1: Structures of HPMC (left) and CNC (right)

Cellulose nanocrystals (CNCs) are nano-scale crystalline cellulose fragments characterised by a high surface-to-volume ratio, alongside inherent biocompatibility and biodegradability[4]. Similar to HPMC, they have emerged as promising excipients in fast-release formulations. Studies have shown that incorporating CNCs can markedly increase the dissolution rate and apparent solubility of hydrophobic drugs compared to conventional excipients. CNC particles adsorb and disperse drug molecules, facilitating the formation of amorphous or metastable solid phases; for example, drug-CNC co-grinding eliminated the drug's crystalline structure (converting it to an amorphous form) and correspondingly enhanced dissolution. CNCs form stable colloidal dispersions in water, which enable uniform API distribution and fast release [5].

This report investigates how key formulation variables affect the properties of the resulting drug-loaded particles. In our study, the candidate polymers (HPMC and CNC) are evaluated, following which we shall investigate the influence of polymer concentration and solvent composition on particle formation. Paracetamol serves as the candidate model drug. Key factors include: (1) **Excipient type** - comparing HPMC versus CNC as the main carrier matrix; (2) **Polymer solution composition** - using purely aqueous solvent vs. a water/ethanol solvent mixture for HPMC to determine how semi-organic solvent influences particle drying and drug dispersion, and (3) **Drug loading** - formulating particles with 5% wt API versus 10% wt API to observe the effect of drug-to-excipient ratio. After producing microspheres via acoustic levitation drying of single droplets, the particles are characterised. **Particle morphology**

is assessed both qualitatively and quantitatively using scanning electron microscopy (SEM) and powder X-ray diffraction (XRD) to evaluate whether the drug remains crystalline or becomes amorphous within the particles. **Solid-state properties** of the drug in each formulation are further examined by Fourier transform infrared spectroscopy (FTIR), which can detect molecular interactions or polymorphic forms. Finally, the **dissolution behavior** of the particles is tested in simulated gastrointestinal fluids to quantify how rapidly the drug is released and dissolved, benchmarking the performance against the ideal of immediate release. All findings are interpreted in light of the overarching project aim: to develop an improved drug delivery platform that provides fast analgesic action. By demonstrating enhanced dissolution and quicker release from our acoustically levitated microparticles, we aim to address the challenges of delayed onset and dosing inconsistency seen with current oral painkillers. Moreover, by using biodegradable excipients instead of non-degradable polymers, the formulated system prioritises patient safety and environmental sustainability.

Methods

Materials

- Hydroxypropyl methyl cellulose (HPMC)
- 2.5% wt CNC solution
- Paracetamol
- Acetone
- Ethanol
- Isopropyl Alcohol
- Methanol

Polymer Selection

Samples of 1.5% CNC and 1.5% HPMC aqueous solutions were prepared by accurately weighing out the required polymer mass and dissolving in deionised water. CNC solutions were pre-prepared and used directly. For HPMC, the hot-cold water dispersion method was employed: the weighed HPMC powder was initially dispersed in one-third of the total volume using hot water (~80 °C) to prevent agglomeration, followed by the addition of cold water (4 °C) under magnetic stirring to achieve the final volume. This two-stage hydration ensured complete dissolution and minimised clumping. Each solution was stirred until a visually homogeneous mixture was obtained before use in levitation experiments.

Preparation of Polymer Solutions

To determine the effect of concentration on the drying kinetics and morphology of the dried particles, 20ml solutions of HPMC in water were prepared with six different concentrations. The concentrations used were 1.25%, 1.5%, 1.75%, 2.5%, 3.25% and 4%. 4% by weight was found to be the upper limit of HPMC solubility in water beyond which HPMC precipitated in the solution even despite mixing overnight. The solutions were

prepared by weighing the required amount of HPMC determined using standard stoichiometric principles and gradually adding them to warm water and mixing using a IKA RO10 magnetic stirrer. The higher concentrated solutions were left overnight on the stirrer to aid in complete mixing.

Acoustic Levitator Setup

The acoustic levitation experiments were conducted using a custom-built levitator enclosed within a humidity-controlled chamber to minimise environmental fluctuations during droplet drying. The system was powered by an RS Pro IPS 303DD DC power supply, which was connected to the levitator's ultrasonic transducers. A Basler acA2000-165um camera was mounted laterally at the levitation node, aligned with the droplet position using iterative adjustment. A diffuse LED backlight was positioned behind the droplet to enable high-contrast imaging. Live imaging was carried out using the Basler Pylon Viewer software, configured to capture one image every 60 seconds throughout the drying process.

Droplet Loading in Acoustic Levitator

To introduce samples into the levitator, an 8 μ L aliquot of each prepared solution was initially drawn using a micropipette. However, for viscous samples (e.g. 2.5 - 4% HPMC and solvent mixtures), a syringe was used for greater control and ease of droplet release. The syringe tip was carefully positioned at the established node location within the standing wave field. An initial voltage of approximately 14 V was then applied to the transducers, generating sufficient acoustic pressure to detach and suspend the droplet. After levitation was achieved, the voltage was gradually reduced to 10.5–12 V to minimise compaction and improve droplet sphericity.

Solvent Screening

The low volatility of water resulted in an extended drying time period within the levitator, limiting the number of particles that could be dried and investigated within the research period, thus acting as a potential bottleneck in the workflow. Therefore, to reduce the drying time required in the levitator, the selected HPMC water solution was further diluted separately with four organic solvents: Acetone, Methanol, Ethanol and IPA. The rationale behind selecting these four solvents was based on their low boiling points and accessibility for research and industrial use. An 80:20 ratio of water to organic solvent was used as this was determined to be the maximum ratio of organic solvent that could be introduced without impacting the solubility of HPMC in the solution. The percentage of HPMC remained constant in all 4 solutions. The samples were then introduced in the levitator using a micro syringe and their drying was investigated using the same voltages as in the case of pure water samples.

Drying Curve Generation

The images captured were analysed using the ImageJ image analysis software. The still images

were generated at an interval of 60s, and the recording was stopped when the droplet size reached a plateau. Based on the volume of particle initially mounted on the levitator node, an equivalent initial diameter was calculated in mm, and this served as the calibration basis in ImageJ. The images were then cropped, processed and analysed within the software to generate a table of particle area against time. Equivalent diameter was then calculated from the area and was normalised using the initial diameter. The drying curve was then generated as a plot of the normalised equivalent diameter against time.

API Solution Preparation

Following selection of a singular HPMC concentration and further dilution with organic solvents, the Paracetamol-HPMC-water-organic solvent solutions were prepared for levitation from encapsulated particles. Two different API loadings are investigated, 5wt% and 10wt%. The paracetamol required was then measured based on stoichiometric principles and mixed with warmed Acetone, Methanol and Ethanol respectively. This solution was then mixed with 2.5% HPMC in water solution. The ratio between water and organic solvent was maintained at 80:20. The final solutions were then mixed using a magnetic stirrer till clear solutions were formed.

Model Fitting

Following the preparation of the API HPMC solution, the drying of the API loaded particles was investigated using the levitator and the images generated were analysed using ImageJ to plot the Equivalent Radius of the particle against time. To quantify the drying kinetics of the particle, the empirical equation Eq X was used to fit the drying data (4). In these τ signifies the characteristic drying time of the process and α signifies a non Fickian parameter (5). Following the fitting of the empirical equation to the experimental dataset, the τ and α are extracted for API solutions with each organic solvent. These and τ and α values are then examined in conjunction with the XRD results and dissolution profiles to quantify the impact of drying kinetics on the particle morphology and subsequent dissolution.

$$R_0 = R_\infty + (R_0 - R_\infty)(1 - \frac{t}{\tau})^\alpha$$

Eqn 1

Particle Morphology

1. Qualitative - Scanning Electron Microscopy (SEM)

The dried particles were then examined under a Scanning Electron Microscope (SEM) to analyse their size and internal morphology. The particles were mounted on a tiny disc using conductive carbon tape. The samples were then coated with Chromium, since it has good conductivity and is finer than gold therefore would not cover the topography of the sample. The disc are then mounted on the SEM Zeiss LEO 1525 plate and

vacuum is created. The aperture, contrast and stigmater is varied between samples to aid in focusing on the particle. The particles are initially analysed with a scanning speed of 2 and then the speed is dropped to 6 to freeze the screen and take snapshots. The particles were analysed with 8 different magnifications to focus on the morphology near the surface, the region between the surface and core and the core.

2. Quantitative – X-Ray Diffraction (XRD)

To assess the solid-state crystallinity of the dried formulations, both control (HPMC-only) and paracetamol-loaded HPMC solutions were analysed via X-ray diffraction. Approximately 7 mL of each prepared solution - comprising either aqueous HPMC or HPMC with 5% and 10% w/w paracetamol - was dispensed into uncovered glass petri dishes using a microsyringe. The solutions were allowed to dry undisturbed for 24 hours in a ventilated fume cupboard, resulting in thin polymer films. Once dried, the films were manually fragmented using a clean scalpel to obtain representative samples for analysis. These fragments were placed onto a flat, zero-background sample holder and gently pressed to ensure uniform surface contact. XRD measurements were performed using a Malvern Panalytical Aeris XRD operating at 40 kV, 15 mA with Cu K α radiation (wavelengths: $\lambda_{K\alpha 1} = 1.5406$ Å, $\lambda_{K\alpha 2} = 1.5444$ Å). Diffraction patterns were collected over a 2θ range of 3°–65° at a scanning rate of 48.2 s. The obtained diffractograms were analysed to identify characteristic peaks of crystalline paracetamol and to assess the degree of amorphous dispersion within the HPMC matrix. The peak intensities at specified angles were then analysed with key drying parameters such as τ and Shrinkage ratio as defined in Eq 2.

$$\text{Shrinkage ratio} = \frac{R_{\infty}}{R_0}$$

Eqn 2

Since both parameters impact API particle mobility and subsequent rearrangement to form a more or less orderly structure, their relationship to the peak intensities in the diffraction pattern of the different formulations were also investigated.

Solid State Interaction Analysis – Fourier Transform Infrared Spectroscopy (FTIR)

FTIR analysis was performed using a Bruker Invenio Platinum ATR (attenuated total reflectance) spectrometer to assess the molecular interactions and solid-state characteristics of the samples. Dried samples - including pure paracetamol crystals, unloaded 2.5% HPMC particles, and 10% paracetamol-loaded HPMC particles (prepared in different solvent systems) were analysed. A small fragment of each sample was placed directly onto the diamond crystal stage of the ATR module. The pressure arm was then lowered and the clamp tightened to ensure consistent contact between the sample and the crystal surface. Spectra were collected using

Bruker's OPUS software, and each scan was background-subtracted and averaged across the wavenumber range of 4000–400 cm⁻¹.

Dissolution

Dissolution behaviour of the dried microparticles was evaluated using time-lapse optical microscopy. Individual particles, including pure paracetamol, unloaded 2.5% w/w HPMC microparticles, and HPMC particles loaded with either 5% or 10% paracetamol were transferred onto clean glass microscope slides. Each particle was then submerged in 100 μ L of distilled water, dispensed using a micropipette. The slides were positioned on the stage of an Olympus BX41M-LED microscope, with magnification adjusted between 4x and 5x as required. Real-time dissolution was recorded using the Basler Pylon Viewer software, which was configured to capture an image every 60 seconds. The image recording was stopped once there is no further dissolution observed. This process was repeated for the selected formulation using simulated gastrointestinal fluids, representing saliva, stomach and intestinal fluids. Time-resolved images were later analysed to assess particle erosion, diffusion front expansion, and relative dissolution kinetics under different pH conditions.

Dissolution Curve Generation

The dissolution image sequences were imported into Fiji (*ImageJ*) for quantitative analysis. Each series was first cropped to isolate the particle and minimise interference from surrounding objects. The threshold setting was adjusted manually to remove background noise arising from dust and illumination variability. Following preprocessing, the images were inverted to enhance contrast between the particle boundary and background. Using the *Analyze Particles* tool, a time-resolved dataset of the projected particle area was obtained. The equivalent circular diameter for each frame was calculated from the measured area, and the corresponding particle volume was derived. Mass as a function of time was then determined by applying the appropriate formulation-specific mixture density (based on HPMC–paracetamol ratios). From the mass–time data, the cumulative percentage mass released was computed and plotted to generate dissolution profiles for each formulation.

Results

1. Drying Curve Analysis (Optical Imaging via Acoustic Levitation)

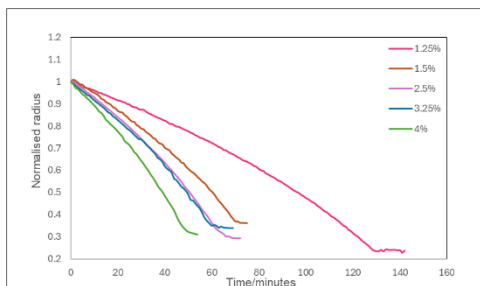


Figure 2: Drying curves(Normalised radius against time) for different HPMC concentrations in water

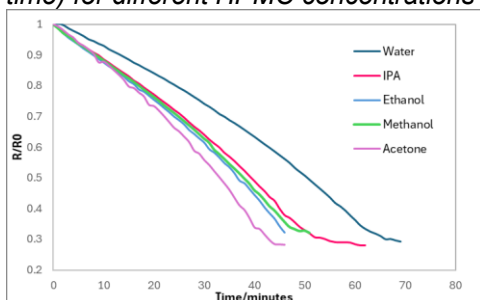


Figure 3: Drying curves for 2.5% HPMC in various organic solvents

Figure 2 shows the graph of the normalised radius of the droplets during levitation against time. Normalised radius is used to account for the inconsistency in initial droplet sizes due to the use of syringe. Figure 3 depicts the normalised radius plotted against time for water and 4 other organic solvents: Ethanol, Methanol, IPA and Acetone. As seen in the figure, dilution with all the organic solvents reduce the drying time needed when compared against water only solution.

2. Model Fitting Analysis

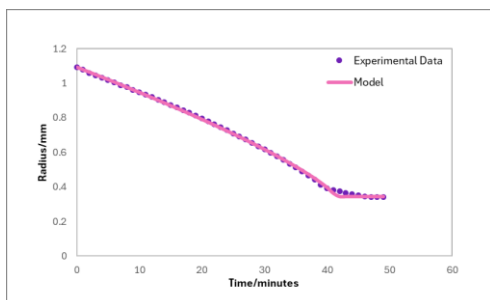


Figure 4: Fitting of the empirical equation to the experimental data for 10% Paracetamol in a solution with 2.5% HPMC, water and Acetone.

Figure 4 represents the fitting between the experimental dataset and the empirical equation. To assess the reproducibility of this fitting the empirical equation with equivalent τ and α parameters is tested against 4 other datasets with the same API-HPMC formulation.

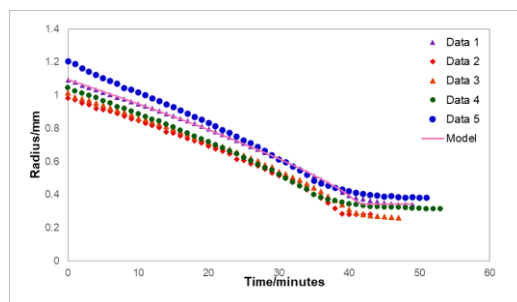


Figure 5: Fitting of the model with 5 datasets

This fitting approach is carried out for API-HPMC solutions in all 3 organic solvents under consideration: Acetone, Methanol and Ethanol and the α and τ values for each fit are summarised in Table 1 in Appendix. For the same formulation, the relationship between the initial and final radius of the particle is outline in Figure 6 and the relationship between τ and the final radius is outlined in Figure 7.

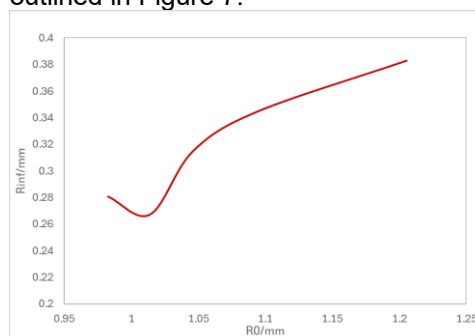


Figure 6: Relationship between initial and final radius of dried particle

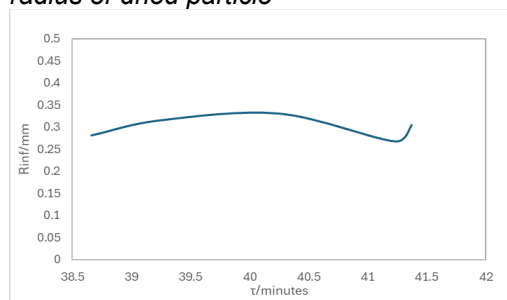


Figure 7: Relationship between final radius of dried and τ

4. XRD

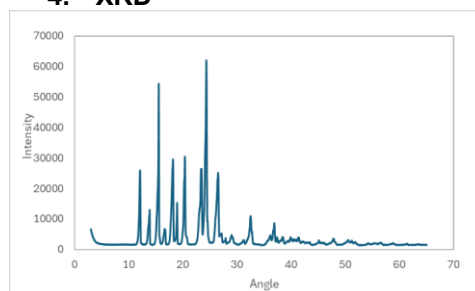


Figure 8: XRD diffraction pattern for pure paracetamol

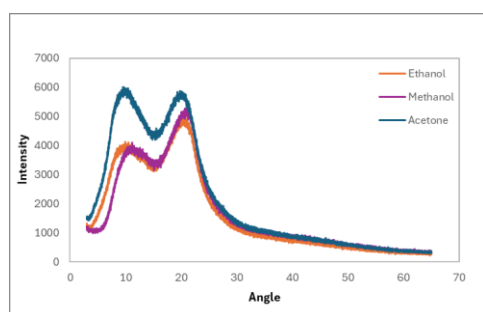


Figure 9: XRD diffraction pattern for formulations with different organic solvents

Figure 8 shows sharp intense peaks at specific angles, predominantly between 10° and 26° approximately. The presence of repeated high intensity sharp peaks confirms that the sample exists predominantly in a crystalline state. Figure 9 shows that drying with these organic solvents result in broad smooth halos instead of sharp peaks with varying peak intensities at specified angles.

To analyse the effect of drying kinetics on the resulting arrangement of Paracetamol in the dried particle, the relationship of τ and Shrinkage ratio to and the peak intensity at 22.34° is analysed as outlined in Fig 10. The rationale behind analysing the intensity at this specific angle corresponds to the presence of a sharp intense peak in the paracetamol sample at this angle and a broad halo in the organic solvent systems centred around this angle.

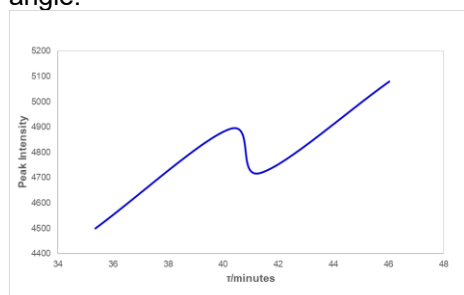


Figure 10: Relationship between τ and peak intensity at an angle of 20.37°

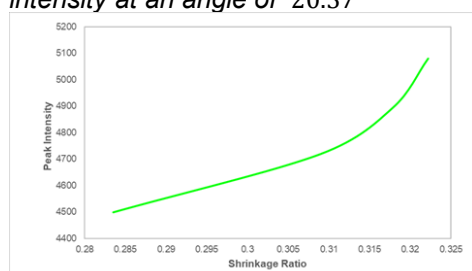


Figure 11: Relationship between Shrinkage ratio and peak intensity at an angle of 20.37°

Dissolution Testing (Optical Microscopy)

The dissolution of particles in aqueous media follows several distinguishable stages that provide insight into the release mechanism. These include initial hydration and swelling, onset of surface erosion or ring formation, and complete dissolution or collapse of the polymer matrix. Optical

microscopy was used to track these morphological changes over time, offering visual confirmation of drug release behaviour and internal structure across different formulations. The evolution of these stages provides a qualitative comparison alongside quantitative mass release profiles.

Table 1: Dissolution images at key stages capture via optical microscopy

		Initial structure (t = 0)	Early Swelling	Onset of Erosion/ Ring Formation	Complete Dissolution
Control	2.5% HPMC (10x)	 t = 0	 t = 1	 t = 3	 t = 12
Dissolution in Water, 10% wt Paracetamol in 2.5% wt HPMC	Acetone (5x)	 t = 0	 t = 1	 t = 5	 t = 26
	Ethanol (5x)	 t = 0	 t = 1	 t = 6	 t = 22
	Methanol (5x)	 t = 0	 t = 1	 t = 4	 t = 34
Dissolution in Simulated GI fluid, 10% wt Paracetamol in 2.5% wt HPMC, Acetone solvent	Intestine (4x)	 t = 0	 t = 1	 t = 3	 t = 10
	Saliva (4x)	 t = 0	 t = 1	 t = 8	 t = 19
	Stomach (4x)	 t = 0	 t = 1	 t = 3	 t = 18

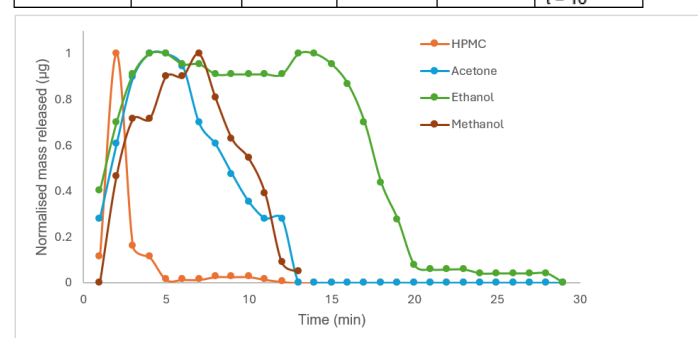


Figure 13: Dissolution of 10% API Loaded HPMC particles in water using various solvents

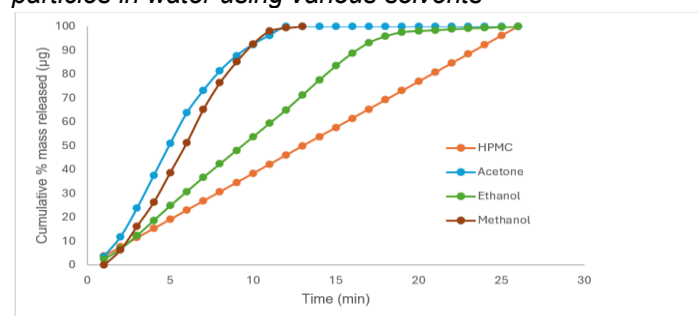


Figure 14: Cumulative % release of 10% API Loaded HPMC particles in water using various solvents

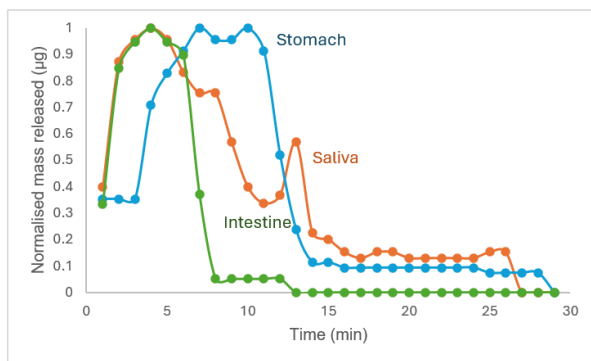


Figure 15: Dissolution of 10% API Loaded HPMC particles in GI Fluids using Acetone as solvent

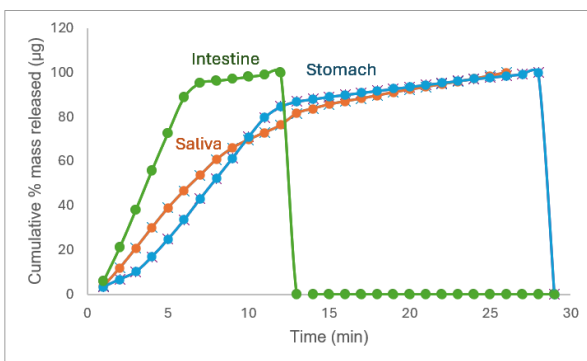


Figure 16: Cumulative % release of 10% API Loaded HPMC particles in GI Fluids using Acetone as solvent

FTIR

Capsules of pure HPMC, paracetamol and various solvent-dried forms were analyzed via FTIR. The resulting spectra are shown in Figure X and X. Table X in Appendix consists of major peak assignments and functional groups used as reference. Table X below summarises the observed peak shifts or broadening for each sample, indicating amorphisation interaction.

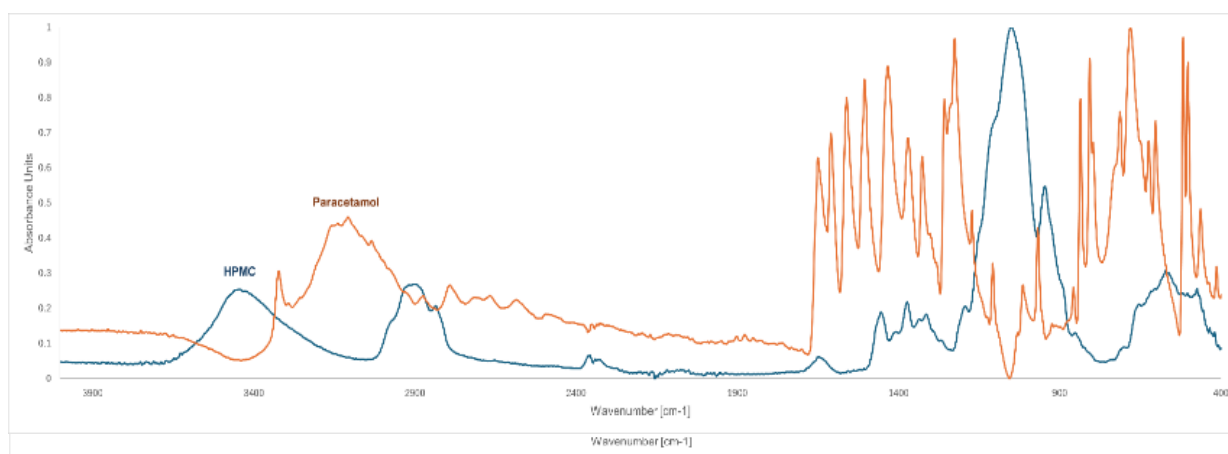


Figure 17: FTIR Spectra for Pure HPMC and Paracetamol (top) and Solvent Systems (bottom)

Table 2: Interpretation FTIR Spectrum for formulations with different organic solvents

Vibrational Band	Pure Paracetamol and HPMC	10% API/HPMC – Acetone	10% API/HPMC – Ethanol	10% API/HPMC – Methanol	Interpretation
O–H/N–H stretching region (broad)	3324 cm ⁻¹ (paracetamol O–H) 3455 cm ⁻¹ (HPMC O–H)	~3400 cm ⁻¹ (broad band)	~3400 cm ⁻¹ (broad, slightly shifted)	~3380–3400 cm ⁻¹ (broad band)	The O–H stretching bands of paracetamol and HPMC merge into a broad peak around 3400 cm ⁻¹ in all formulations, indicating hydrogen bonding and amorphous dispersion.
C=O stretching (amide I of paracetamol)	1656 cm ⁻¹ (sharp) No C=O in pure HPMC	~1654 cm ⁻¹ (distinct peak)	~1648 cm ⁻¹ (slightly broadened, shifted)	~1645 cm ⁻¹ (broad, slight red-shift)	The C=O band in ethanol and methanol samples shifts ~8–11 cm ⁻¹ lower and broadens, consistent with stronger drug–polymer interactions. In acetone, the band remains sharp and near its original position, suggesting partial crystallinity
N–H bending (amide II of paracetamol)	1565 cm ⁻¹ (distinct) No corresponding band in HPMC	~1562 cm ⁻¹ (visible, somewhat sharp)	~1558 cm ⁻¹ (broadened)	~1555 cm ⁻¹ (broadened)	The amide II band (~1560 cm ⁻¹) also shifts and broadens in alcohol-based particles, further supporting amorphisation via hydrogen bonding.
Aromatic ring C=C stretch	1609 cm ⁻¹ (paracetamol) ~1605 cm ⁻¹ (minor in HPMC)	~1608–1610 cm ⁻¹ (sharp)	~1605 cm ⁻¹ (slightly reduced intensity)	~1605 cm ⁻¹ (slightly reduced intensity)	The aromatic C=C stretch (~1605–1610 cm ⁻¹) remains unchanged, indicating no chemical modification of paracetamol.
–CH ₃ /2 bending vibrations	1445 cm ⁻¹ (paracetamol C–C / CH bending) 1452 cm ⁻¹ (HPMC CH ₂ bend)	~1440–1450 cm ⁻¹ (broad)	~1440 cm ⁻¹ (broad)	~1440 cm ⁻¹ (broad)	CH bending bands (1440–1450 cm ⁻¹) broaden in all samples, showing physical mixing and confirming no new bond formation.

Discussion

Amorphous Drug Form and Dissolution Kinetics

Amorphous behavior is crucial in drug-excipient formulations given that the amorphous form of a drug generally exhibits higher apparent solubility and faster dissolution rates than its crystalline counterpart [5]. This is attributed to the higher free energy and lack of a rigid crystal lattice in the amorphous state, which means less energy is required for drug molecules to escape into solution. Hence, an amorphous drug would be able to dissolve more rapidly and to a greater extent, potentially enhancing oral absorption and bioavailability.

Although paracetamol has moderate aqueous solubility, converting some or all of it to an amorphous form via capsule formulation accelerates its release. For example, spray-dried paracetamol-polymer dispersions show significantly faster dissolution compared to its pure crystalline form: one study reported a mean dissolution time of only ~2.8 minutes for spray-dried amorphous paracetamol/PVP particles, compared to ~19.7 minutes for intact crystalline paracetamol[6]. The dramatic enhancement in dissolution was attributed primarily to reduced crystallinity and the presence of amorphous drug domains. In such amorphous solid dispersions, the polymer excipient (HPMC) not only serves as a carrier but also acts as a stabiliser that inhibits recrystallisation, maintaining the drug in a disordered, rapidly dissolving state. The goal of our experimentation is then to keep paracetamol at least partially amorphous and well-dispersed at the molecular level via integration with HPMC-excipient. Ultimately, this would facilitate quicker drug release in the gastrointestinal environment, which is beneficial for achieving rapid analgesic action in oral delivery.

Polymer Matrix Selection (HPMC vs CNC)

HPMC was selected as the primary excipient instead of CNC because it demonstrated far more favourable structural and solid-state characteristics for forming stable, fast-dissolving microcapsules. SEM imaging showed that HPMC formed smooth, continuous matrices upon drying, whereas CNC droplets dried into irregular, brittle aggregates with poor morphological integrity. XRD confirmed this distinction: CNC retained strong crystalline peaks, indicating a rigid ordered structure that would encourage paracetamol recrystallisation and slow dissolution.

Solvent Effects and Drying Kinetics

Drying Curve Analysis

All the HPMC concentrations used for drying inside the levitator depict the same drying behaviour. Initially the solvent concentration is large enough and therefore evaporation of the solvent dominates the drying rate; this can be observed by the approximately linear section of the drying profile. As the solvent inside the particle gets evaporated, the polymer particles structure themselves inside the particle. The drying is then dominated by the diffusion of the solvent through the gaps between these polymer particles. As expected, with the increase of the polymer concentration the drying time needed decreases. A HPMC concentration of 2.5 wt% is selected for further investigation as it had one of the lowest drying times and its viscosity and consistency enabled easier loading of the particle onto the levitator and consistent initial droplet sizes. As shown in Figure 3, dilution with organic solvents reduced the drying time needed as well as preserved the drying behaviour. IPA was not considered for further investigation as it had the least improvement in drying time needed. Ethanol, Methanol and Acetone were all considered for further investigations as they had varying solubility with HPMC and Paracetamol.

1. Model Fitting Analysis

The parameters used in the empirical equation used to fit to the experimental data has two key components that provide insights into the drying kinetics; τ is the characteristic drying time of the droplet, a larger value of τ indicates slower drying time overall. The α parameter is a non Fickian parameter. As seen in Fig 5, when the model is tested with 5 different datasets, the empirical model fits reasonably well. It can be seen that while there is a strong positive correlation between R_∞ and R_0 , τ has an extremely small impact on R_∞ . The strong positive correlation between R_0 and polymer concentration to R_∞ can be attributed to the fact that the total solid content in the droplet scales linearly with both R_0 and concentration and since both HPMC and Paracetamol are non-volatile, they are retained in the dried particle therefore a higher total solid content would result in a larger R_∞ .

Structural Characterisation

XRD

The diffraction pattern observed in an XRD spectrum can be attributed to the constructive interference of scattered X rays reflected off the atoms within the sample. The constructive interference produces the distinct diffraction pattern that is characteristic of the sample. In crystalline materials, the atoms are arranged in regularly ordered lattices therefore when X rays scatter from them, they interfere coherently following Bragg's law hence form sharp intense

peaks at specific angles. In amorphous material, atoms lack the orderly lattice arrangements hence the scattered X rays interfere inconsequentially and result in broad halos rather than sharp peaks. As seen in Figure 8 and Figure 9, paracetamol has the characteristic diffraction pattern of a crystalline material and when it is combined with the HPMC solvent system, the system is amorphous in nature. To outline the impact of drying kinetics on the resulting nature of the particle, the relationship between τ and peak intensity at the specified angle is analysed. It is observed that with an increasing τ the peak intensity also increases. An increase in peak intensity was interpreted as a higher degree of crystallinity in the sample and greater local ordering. This positive correlation be attributed to the fact a higher τ corresponds to a larger drying time therefore allows higher particle mobility to form the ordered structures.

FTIR

The extent of peak shifts/broadening varies slightly with the solvent used to prepare the particles, suggesting differences in the extent of integration into the polymer matrix. In the drug-loaded formulations, the paracetamol peaks are significantly broader and slightly shifted compared to the pure crystalline form. For example, the phenolic O–H/N–H stretch (which is a sharp band in pure paracetamol around 3325 cm^{-1}), becomes a very broad, merged band with HPMC's O–H stretch around $\sim 3400\text{ cm}^{-1}$ in the composites. Similarly, the paracetamol amide C=O stretch (normally $\sim 1655\text{ cm}^{-1}$) appears at a somewhat lower wavenumber ($\sim 1640\text{--}1650\text{ cm}^{-1}$) and with broad profile in the loaded particles. This broadening and slight red-shift of hydrogen-bond sensitive bands (O–H, N–H, C=O) is strong evidence that paracetamol is present in an amorphous form within the HPMC matrix, with extensive drug–polymer hydrogen bonding interactions. The hydrogen-bond formation between paracetamol and the hydroxyl-rich HPMC leads to shorter intermolecular distances and hence lower vibrational frequencies and wider absorption bands. In practical terms, these spectral features indicate that the drug is molecularly dispersed in the polymer (forming an amorphous solid dispersion) rather than remaining as separate crystals.

Moreover, the FTIR spectra do not show any new absorption bands or any disappearance of HPMC's own peaks, which indicates that the HPMC structure remains intact and there is no chemical breakdown of either component. The drug-

excipient interaction is predominantly through hydrogen bonding and physical mixing. This finding is consistent with expectations for a stable drug–polymer system: HPMC is generally inert and interacts via intermolecular forces. The presence of all key paracetamol peaks (albeit shifted/broadened) in the spectra of the microparticles, combined with the broad O–H band of HPMC, suggests the formulation is a physical mixture at the molecular level (amorphous solid dispersion) rather than forming any new covalent species. Such interactions can improve the drug's amorphous stability and are desirable for ensuring the drug remains dispersed; concurrently, the lack of chemical shifts in unrelated regions means the drug's chemical identity is preserved[7]

Water-Based Dissolution and Solvent Selection

When comparing the dissolution of dried particles prepared in different solvents, notable differences in performance were observed. The control HPMC-only particle displayed gradual swelling followed by slow erosion, reflecting its gel-like matrix that resists rapid diffusion. In contrast, drug-loaded particles prepared using organic solvents showed significantly faster dissolution and sharper ring breakdown. Among these, the acetone-based formulation exhibited the fastest release profile and completed dissolution, as supported by both the normalised mass (Figure 13) and cumulative release plots (Figure 14). The rapid drop in particle size and increase in cumulative drug release suggests a highly amorphous structure, likely formed due to the fast evaporation rate of acetone during levitation.

This behaviour was confirmed via FTIR (hydrogen bonding and C=O shift, and SEM (non-crystalline morphology), supporting the formation of a molecularly dispersed amorphous solid dispersion. Ethanol and methanol composites also facilitated reasonably fast dissolution but showed slower onset compared to acetone. Moreover, methanol-derived particles displayed initial swelling resistance, while ethanol formulations had residual crystalline structures, indicated by incomplete ring collapse and gradual plateauing. These findings validate **acetone as the most suitable solvent** for encapsulation in our study.

Gastrointestinal Fluid Dissolution: Evaluating Site-Specific Release

Using the acetone-based formulation (10% paracetamol in 2.5% HPMC), dissolution was evaluated in simulated gastrointestinal fluids: saliva, gastric fluid (stomach), and intestinal fluid. Compared to the neat HPMC control, the drug-loaded particles demonstrated superior dissolution performance in all media. In simulated saliva, the formulation showed moderate swelling followed by surface erosion and consistent mass release. The

stomach simulation, however, yielded the fastest and most complete dissolution, with rapid hydration and collapse of the matrix within ~12 minutes, supporting the formulation's ability to deliver paracetamol efficiently at this site. In intestinal fluid, the particles displayed lower overall release and prolonged integrity, suggesting reduced solubility of the encapsulated drug at this pH. This is advantageous for site-specific targeting,

as the formulation appears preferentially released in gastric and oral environments, where paracetamol is intended to act.

Conclusion & Outlook

This study demonstrated the feasibility of using acoustic levitation as a precision tool for engineering biodegradable polymer-drug microcapsules with enhanced dissolution performance for paracetamol delivery. By systematically evaluating excipient choice, solvent system, drying kinetics, and API loading, HPMC was identified as the most suitable biodegradable carrier due to its ability to stabilise paracetamol in a predominantly amorphous state. In contrast, cellulose nanocrystals produced crystalline composites and were therefore unsuitable for rapid-release formulations.

Drying curves for API-loaded droplets were modelled using an empirical equation, yielding the parameters τ and α . Correlating these with solid-state properties showed that slower drying (larger τ) and greater shrinkage promoted higher XRD peak intensities, indicating increased crystallinity. Acetone was determined as the solvent of choice following characterisation. Water-based dissolution tests further validated this: acetone-dried particles exhibited the fastest and most complete release, consistent with their amorphous microstructure and reduced diffusion resistance. Dissolution in simulated GI media showed site-dependent release profiles, with rapid and near-complete dissolution in gastric and salivary fluids.

Future work should focus on enhancing scalability and pharmaceutical translatability. First, optimisation of the HPMC-API ratio and solvent composition could further tune the balance between amorphous stability and mechanical integrity. Since acoustic levitation models single-droplet spray drying, validating trends at pilot-scale spray drying would confirm industrial relevance. Additionally, in-depth thermal and stability studies (DSC, humidity stress, recrystallisation kinetics) are required to understand long-term amorphous stability, which remains a major challenge for paracetamol formulations.

Expanding the platform to other analgesics or poorly soluble APIs would test the applicability of the method and the suitability of HPMC as a universal biodegradable carrier. Finally, integrating

mechanistic modelling of dissolution and GI transit could support prediction-based capsule design, moving toward personalised fast-acting analgesic delivery systems.

References

- [1] Y. Yue *et al.*, "Efficacy and speed of onset of pain relief of fast-dissolving paracetamol on postsurgical dental pain: Two randomized, single-dose, double-blind, placebo-controlled clinical studies," *Clin Ther*, vol. 35, no. 9, pp. 1306–1320, Sep. 2013, doi: 10.1016/j.clinthera.2013.07.422.
- [2] N. Sharma and A. Mehta, "Role of Excipients in Drug Formulation and Stability Enhancement: Strategies for Optimized Therapeutics".
- [3] C. E. Udoh, V. Garbin, and T. Cabral, "Downloaded via 146.179.166.100 on December 11," vol. 15, p. 9, 2025, doi: 10.1021/acs.langmuir.6b01799.
- [4] R. Nath, N. Das, and L. Pachau, "CELLULOSE NANOCRYSTAL-INCORPORATED CO-PROCESSED EXCIPIENT IN TABLET FORMULATION," *CELLULOSE CHEMISTRY AND TECHNOLOGY Cellulose Chem. Technol*, vol. 58, no. 2, pp. 31–43, 2024.
- [5] E. Leyk and M. Wesolowski, "The effect of cellulose derivatives on paracetamol crystallinity reduction," *Journal of Thermal Analysis and Calorimetry 2022 147:18*, vol. 147, no. 18, pp. 10037–10048, Apr. 2022, doi: 10.1007/S10973-022-11312-9.
- [6] F. Sadeghi, M. Torab, M. Khattab, A. Homayouni, H. Afrasiabi Garekani, and A. H. Garekani, "Improvement of Physico-mechanical Properties of Partially Amorphous Acetaminophen Developed from Hydroalcoholic Solution Using Spray Drying Technique," *Iran J Basic Med Sci*, vol. 16, pp. 1100–1108, 2013.
- [7] E. Leyk *et al.*, "Stability of Paracetamol Amorphized by Co-Melting with Various Cellulose Derivatives," *Applied Sciences 2024, Vol. 14*, vol. 14, no. 15, Jul. 2024, doi: 10.3390/APP14156709.

Defining Structural Nateness For De Novo Antibody Design

Angelina Jia and Alexia Thill

Department of Chemical Engineering, Imperial College London, U.K.

Abstract This study establishes a quantitative definition of *structural nateness* for bound antibodies and nanobodies and applies it to evaluate diffusion-based backbone generative models. Native structural datasets were curated from high-quality SABDab entries, filtered for resolution, completeness, and the post-training publication date to avoid data leakage. Computationally generated CDR backbones from RFantibody and RFD3 were compared with these native structures using two complementary metrics: root-mean-square deviation (RMSD) of α -carbon coordinates and mean dihedral distance (MDD) derived from ϕ/ψ torsion angles. Across CDR1 and CDR2, native test structures consistently exhibited substantially lower RMSD and MDD values than RFantibody, indicating closer alignment with experimentally observed conformational space. RFantibody frequently generated non-native CDR length combinations and sampled sterically disallowed regions of Ramachandran space. For CDR3, dataset separation was reduced, reflecting its greater conformational diversity, but RFD3, a newer model trained at all-atom resolution, produced structures more similar to the native distribution than RFantibody. Correlation analyses demonstrated that RMSD and MDD capture distinct geometric features and often identify different closest native matches. Together, these results show that structural nateness can be quantified using complementary geometric metrics and provide a principled basis for evaluating generated CDR backbones. The framework developed here offers a foundation for improved nateness assessment and future integration into de novo antibody and nanobody design pipelines.

1. Introduction

Antibodies are key players in the immune system, functioning as highly specific proteins able to identify and neutralise antigens with precision. Their Y-shaped structure and modular variable domains allow them to recognise a broad range of molecular targets while limiting unintended interactions. High specificity, strong affinity, and generally low immunogenicity have made antibodies essential to modern biotherapeutics, with applications ranging from cancer immunotherapy to modulating autoimmune diseases¹. Because they enable customised treatments, antibody therapies are experiencing rapid growth in the biopharmaceutical industry, with global revenues projected to hit 1 trillion USD over the next ten years^{1,2}. Despite the field's potential, antibody development for targeted therapies still has substantial room for improvement. Extensive efforts are being made to develop de novo antibodies with optimised properties, such as enhanced specificity for a chosen target.

Structurally, conventional antibodies comprise two identical heavy chains and two identical light chains, which together form the crystallisable fragment (Fc) and antigen-binding fragment (Fab) (see Figure 1). The Fab contains the variable domains of the heavy (VH) and light (VL) chains, which together make up the Fv region, where antigen recognition and binding occur. Antibody diversity arises from V(D)J recombination, generating vast sequence and structural variation within a biophysically constrained design space^{3, 4, 5}.

Within the Fv region, binding specificity is determined by six hypervariable loops, known as complementary determining regions (CDRs), with three from each chain. The CDR3 loop shows the greatest structure and sequence variability. These loops form the paratopes that interact directly with antigen-binding sites or epitopes^{6, 7}.

Nanobodies are single-domain antibodies (VHHs) produced by camelids and some shark species^{5, 8}. Despite their reduced size, they retain high specificity and affinity while exhibiting comparable thermodynamic stability and can also reach epitopes typically inaccessible to

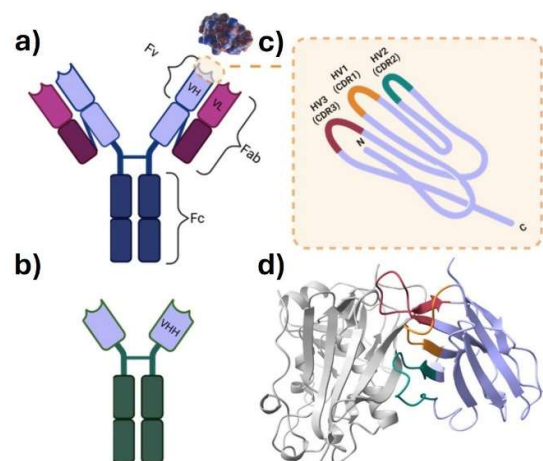


Figure 1. a) Structural overview of conventional and b) camelid heavy-chain antibodies, c) CDR regions, and d) ChimeraX rendition of a sample nanobody, where the full antibody-antigen complex can be seen, with the CDR regions highlighted in their corresponding colours.

conventional antibodies. These properties have driven growing interest in their therapeutic potential, particularly for achieving deeper tissue penetration and more effective engagement with active sites^{5, 8}.

Antibody discovery has traditionally depended on in vivo immunisation or in vitro library screening, both time-consuming, resource-intensive approaches with ethical and practical limitations^{5, 6}. In silico methods, which apply computational pipelines to design and optimise antibody candidates, provide faster, scalable alternatives that lessen dependence on early experimental screening. Unlike traditional methods, in silico workflows could enable more precise antibody engineering targeting predefined epitopes^{4, 5, 9}. This would be a critical advantage in neutralising rapidly evolving viruses such as SARS-CoV-2, as initial designs could focus on epitopes less prone to mutation⁴. Consequently, computational antibody development presents a promising and dynamic avenue for

producing scalable, precise, and cost-effective monoclonal antibodies.

A persistent challenge in the computational design and assessment of antibodies, however, is determining whether a newly generated sequence or structure possesses properties conducive to successful therapeutic development. Therapeutic antibodies must also withstand formulation stresses their *in vivo* counterparts do not face, including temperature fluctuations, long shelf-life, and high-concentration conditions⁹. Rather than estimating each property through simulation or empirical models, which remain computationally prohibitive or unreliable due to a lack of standardised, comparable property data, one can bypass direct estimation and design antibodies that resemble those already successful in biological and clinical settings^{6, 10}. This pattern-recognition principle underlies nativeness.

Although computational methods have improved, *in vivo* immunisation remains the gold standard, producing antibodies with inherently favourable developability traits: long half-life, low immunogenicity, and minimal toxicity¹⁰. Naturally optimized antibodies can guide *in silico* design, focusing candidate selection on sequences more likely to succeed downstream. Traditionally, nativeness has been defined by the similarity of a sequence to the distribution of naturally occurring, immune-derived antibodies.

However, nativeness defined by structure remains relatively underexplored. While sequence-based metrics such as AbNativ and Sapiens exist, they do not directly account for three-dimensional conformations or loop geometries¹⁰. In this work, we define nativeness as a measure of how closely a candidate antibody or nanobody aligns with structural motifs found in experimentally validated structures. Unlike sequence-based approaches, this definition focuses solely on solved structures and distinguishes only between antibodies and nanobodies, without relying on species-specific or explicitly immune-derived datasets.

This work introduces a quantifiable definition of structural nativeness, enabling an initial assessment of generated antibody backbones independently of sequence-based metrics. By comparing dihedral-angle patterns and spatial differences between crystallized and diffusion-generated antibodies against a curated crystallised reference set, we aim to:

- Identify clear distinctions between computationally generated and experimentally derived structures, thereby defining structural nativeness.
- Benchmark RFantibody and RFD3 on their ability to generate plausible antibody structures.
- Create a structural analysis pipeline, comparing structural information of a query set of antibodies to a native reference set.

We thereby establish the foundation for a structural nativeness scoring tool which can be further developed for integration into antibody and nanobody design workflows.

2. Background

To generate new antibody candidates *in silico*, a standard design pipeline is typically followed. After identifying epitopes on a target antigen and selecting a set framework,

thousands of CDRs are generated to form a new Fv region. The CDR loops are subsequently grafted onto a human or native framework¹¹. This generation is possible through diffusion-based deep learning methods such as RFantibody¹², and RFDiffusion3 (RFD3)¹³, developed by the Baker Lab, a leader in the protein diffusion field, or even Germinal, or Boltzgen^{14, 15}.

The obtained backbone, which uses glycine as a placeholder due to its lack of side chain, is exempt from sequence information. Inverse folding is therefore applied, associating a plausible amino acid sequence to the bare structure. This is done using software such as ESM-IF1 or ProteinMPNN¹⁶.

Following the sequence generation for a given backbone, structure predictors can be used for self-consistency, verifying if the designed sequences indeed adopt the correct structure and form the intended antibody-antigen interactions¹⁷. To select variants with high developability profiles⁶, candidate antibodies or nanobodies are filtered and scored on a wide variety of criteria such as solubility, stability, and binding affinity via computational scoring tools and interface analysers such as CamSol, AbMelt, and NanoMelt^{6, 18, 19}. Post-filtering, the most promising candidates are synthesised and subjected to experimental tests using high-throughput screening or biophysical assays, further assessing binding abilities and properties¹².

Although attractive for its target tuning possibilities, the field has yet to consistently demonstrate the production of developable *de novo* antibodies, with only a few occurrences of *in silico*-designed antibodies and nanobodies expressing and showing similar binding affinities as *in vivo* or *in vitro* synthesised proteins^{14, 15}.

2.1 Nativeness: a developability shortcut?

In the context of *de novo* antibody design, one can define nativeness as the degree of resemblance of a given candidate to a native distribution of antibodies. It is hoped that more native-like candidates will have properties similar to existing antibodies and offer a robust baseline for further optimisation, bypassing complex property optimisation processes which remain computationally prohibitive or unreliable due to a lack of standardised, comparable property data⁶ and avoiding trade-offs between expressibility and pharmaceutical synthesis-oriented targets^{5,7,9}. For example, antibodies engineered for higher affinities are often associated with increased aggregation and viscosity, undesirable for pharmaceutical applications⁶.

Efforts to distinguish generated antibodies and nanobodies from native ones have largely focused on sequence nativeness. Several tools can now compare real and computationally generated protein chains to evaluate how “native-like” their amino acid sequences are. The leading method in this area is AbNativ which scores nativeness across both framework and CDR regions¹⁰. This enables reliable identification of native and non-native candidates in *de novo*-designed antibodies, capturing humanness, as well as in nanobodies, where it evaluates camelid nativeness.

Notably, nanobodies are also assessed for humanness, and the most promising candidates are selected through a Pareto-style optimisation that balances both VHH

nativeness and humanness. Similarly, Hu-mAb is a classifier that evaluates the humanness of antibody sequences and suggests framework mutations to enhance developability²⁰. The objectives of sequence nativeness evaluation are therefore to select promising candidates for development and to enable the humanisation of generated sequences. Sequence nativeness implicitly enhances important developability characteristics, including low immunogenicity, solubility, and stability, by capturing evolutionary constraints that have shaped antibody sequences. It can also function as a safeguard during subsequent property-optimisation steps.

Beyond sequence-level assessment, structural nativeness could become a valuable early candidate-selection tool in in-silico antibody design. The structural features of generated antibodies, particularly local CDR motifs and loop conformations, directly influence key developability attributes, including binding behaviour and conformational stability^{6, 21}.

Although structural data is far scarcer than sequence data (thousands of solved structures versus millions of sequences^{22, 23}) due to the experimental challenges of protein crystallography²⁴, incorporating structural nativeness can still provide meaningful guidance. By evaluating the plausibility of generated backbones early in the design pipeline, structure-based nativeness metrics may help steer models toward more realistic

conformations, reduce downstream computational costs, and yield antibodies whose binding properties more closely resemble native ones, ultimately supporting more efficient optimisation.

Various design tools integrate structural parameters of antibodies, both for structure search and humanisation purposes^{3, 25}. Among these, the predominant tools are those based on query-driven search.

Given an input structure, software such as PyIgClassify or MASTER identify the best matches, that is, the closest structures based on defined structural comparison criterion. While MASTER focuses on residue distances between compared antibodies, PyIgClassify is based on dihedral angle comparison^{7, 25}. Dihedral angles describe the rotation between two planes defined by atoms bonded to adjacent carbon atoms.

Drawing inspiration from structural comparison approaches, this work investigates a range of geometric variables to formulate a definition of structural nativeness.

3. Results

Recent advances in in silico protein design now enable the generation of antibody CDR backbones that can be grafted onto a specified framework region. To evaluate the ability of these models to produce plausible antibody structures and potentially guide de novo antibody design, we compared structural metrics extracted from generated CDRs, namely dihedral angles and α -carbon coordinates, with corresponding metrics from native antibody structures obtained from The Structural Antibody Database (SAbDab).

Although both antibodies and nanobodies were investigated (see section 4 of Supplementary Materials), this report focuses on the elucidation of the structural attributes of the latter.

3.1 Data parsing, obtaining native and generated sets

To benchmark diffusion models, we first obtained native antibodies and nanobodies structural data. From the initial 4,573 solved experimental antibody complexes and 1,113 nanobody complexes catalogued in SAbDab²², we constructed native test and reference datasets, filtering for crystallographic resolution, residue completeness, and the exclusion of unbound complexes to focus on static, binding-competent conformations, and avoid complex dynamic modelling. Structures with exceptionally long or unusually complex CDR loops were also excluded to prevent the over-representation of rare structural features. The test subsets were restricted to structures solved and published after 1 December 2024, the release date of RFD3, to prevent bias, given that RFD3 was trained on SAbDab. Additional filtering ensured diversity in antigen types and antibody sequences.

Because ϕ/ψ torsion angles and RMSD values are computed directly from atomic coordinates, crystallographic structures inherit uncertainty present in the underlying refined model. Backbone atoms typically exhibit positional errors of a few tenths of an ångström even in well-refined structures, meaning that models at or below our 3.5 Å resolution cutoff still introduce small but non-negligible variation into the resulting dihedral angles and distance-based metrics. Applying this resolution filter, together with excluding structures missing key backbone

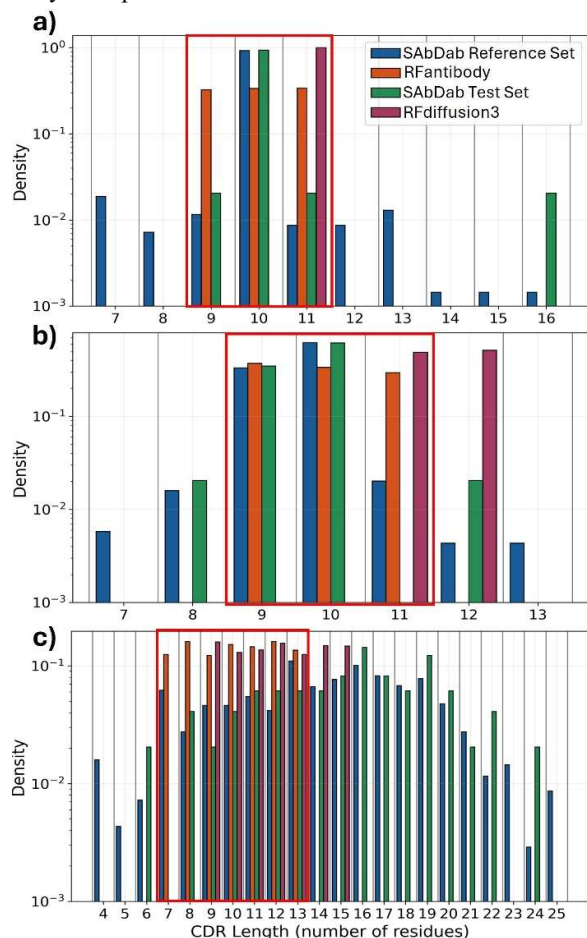


Figure 2. CDR Lengths distribution across native and generated sets for a) CDR1 b) CDR2 and c) CDR3, with the final length range used for analysis framed in red.

or side-chain atoms in the residues of interest, therefore reduces, but cannot fully remove, measurement error within the native structural datasets²⁶.

We then retrieved epitopes for all antigens in the test set and generated candidate antibody structures using two protein diffusion tools from the Baker Lab. The first tool, RFantibody, represented the original standard in antibody diffusion, and RFD3, its more recent, all-atom based version. CDR backbones were generated on specified frameworks, covering the length ranges indicated in Figure 2

After applying the filtering criteria, the reference set contained 2,680 antibodies and 693 nanobodies, and the test set contained 69 antibodies and 49 nanobodies. RFantibody generated 566 antibody and 442 nanobody backbones, and RFD3 produced 2,000 nanobody backbones.

For RFD3, the generated structures were provided directly by the Baker team; only VHH backbones were produced, and we had no control over the CDR length ranges or generation settings.

3.2 Conformational antibody profiling using Ramachandran space

Initially, after calculating the dihedral angles for all residues in each dataset, Ramachandran plots were generated to assess whether qualitative differences in the backbone angles could be observed between native and generated structures. As seen on the Ramachandra plots for CDR1 in the nanobodies (see Figure 3), the test set angles fall largely within the same regions occupied by the reference structures. However, the RFantibody set shows clear deviations, sterically disallowed backbone conformations in CDR1 and lower ψ values than typically found around the left-handed α helix conformation (see Supplementary Figures 3.5a) (CDR2) and 3.5b) (CDR3))²⁷. These deviations suggest the model does not consistently generate backbones that are within steric

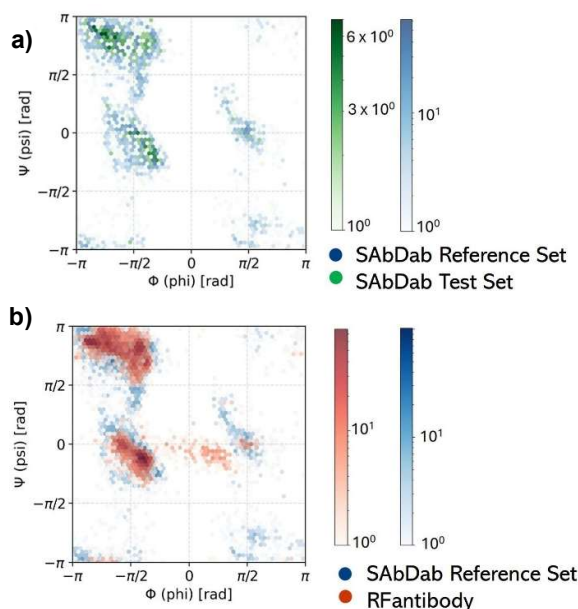


Figure 3. Ramachandran plots for CDR1. **a)** SAbDab reference (blue) and SAbDab test (green) datasets showing the native conformational space. **b)** Comparison of SAbDab reference (blue) and RFantibody-

allowances. Interestingly, even in regions where RFantibody and the test sets overlap with the reference space, their distributions concentrate in slightly different areas.

These visual contrasts between a native and generated set motivated a deeper assessment of those deviations, via the quantitative metrics developed in this work.

3.3 Structural comparison of nanobody datasets

To ensure a fair comparison, CDR length ranges were restricted to those present in the RFantibody set, as the primary goal was to quantify differences between the test and RFantibody datasets. RMSD and MDD values were computed for each query structure against all reference structures in CDR 1, 2, and 3. For each individual query structure, the minimum RMSD and MDD values found within the reference set were retained as the closest structural match. The resulting distributions are shown in Figure 4.

For CDR1, the test set exhibits a significantly lower mean MDD (0.177) and RMSD (0.939 Å) values than RFantibody (0.263 and 1.77 Å), respectively, indicating that test structures have closer local backbone agreement with the reference nanobody set. In CDR2, a similar pattern arises. The mean MDD and RMSD values for RFantibody (0.268 and 1.83 Å) remain comparable in magnitude to those in CDR1, whereas the test set again moves closer to the reference set, with substantially lower values (0.069 and 0.610 Å). This results in a larger separation between the test and RFantibody distributions than observed in CDR1. These differences in distribution were also found to be statistically significant under a two-tailed t- test (Supplementary Table 2.2).

While RFD3 structures were not considered in the comparison against CDRs 1 and 2 due to the unintended out-of-distribution lengths range generated, the CDR3 lengths generated remained within a range suitable for comparison. The test, RFantibody, and RFD3 sets

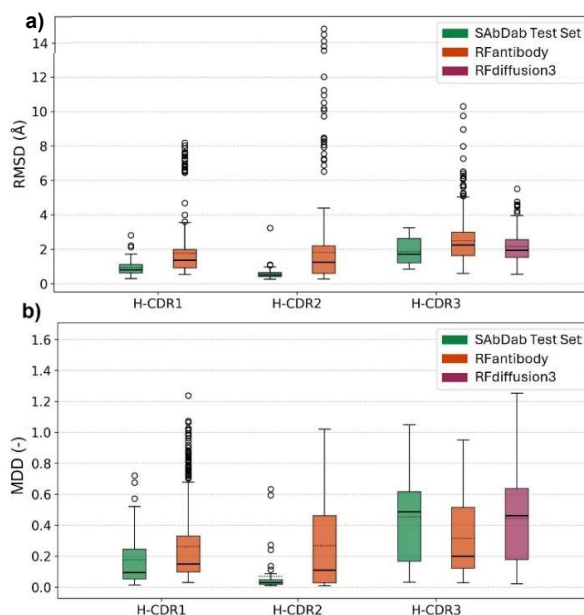


Figure 4. a) RMSD and **b)** MDD Distributions across native and generated sets for closest reference match. Dotted lines represent average values, solid line median values.

showed mean RMSDs of 1.86, 2.49, and 2.17 Å, consistent with the pattern of other CDRs, where RFD3 lies in between the two. In contrast, however, the mean MDD values (0.454, 0.315, and 0.444), did not share this trend. The first observation is that RFantibody appears to have a lower mean value, this was however, not found to be statistically significant. Even so, the test set distribution does not reach lower MDD values observed in RFantibody (Figure 4b). Notably, although the test and RFD3 datasets lie farther from the reference distribution than RFantibody, they exhibit nearly identical mean MDD values, with RFD3 showing a slightly broader spread. In fact, for both RMSD and MDD metrics, the RFD3 distributions are consistently closer to those of the test set than to RFantibody for CDR3.

Because RMSD and MDD values were computed for each CDR independently, we also recorded the CDR1 and CDR2 lengths to examine how length combinations were distributed across datasets. The RFantibody set included several CDR1-CDR2 length pairs absent from the native datasets, such as (9,9) and (11,11), while showing limited representation of the combinations most common in native antibodies. In contrast, both native datasets were heavily concentrated in two length combinations (10,10) and (10,9), which together account for 89.8% of structures in the test set and 88.8% in the reference (Supplementary Figure 3.1).

3.4 Correlation between positional and angular metrics

To assess the relevance of using positional and angular data conjointly, correlation analyses were performed on both the test and RFantibody datasets.

Table 1. RMSD-MDD Closest Matches Correlation Coefficients. S and P are used to indicate Spearman and Pearson coefficient.

	CDR1		CDR2		CDR3	
Coefficient	S	P	S	P	S	P
SAbDab Test Set	0.76	0.698	0.573	0.717	0.815	0.754
RFantibody Set	0.65	0.603	0.891	0.546	0.538	0.465

In most cases, MDD and RMSD closest match values seem positively correlated, as shown in Figure 5 and confirmed by the Spearman (S) and Pearson (P) coefficients shown in Table 1.

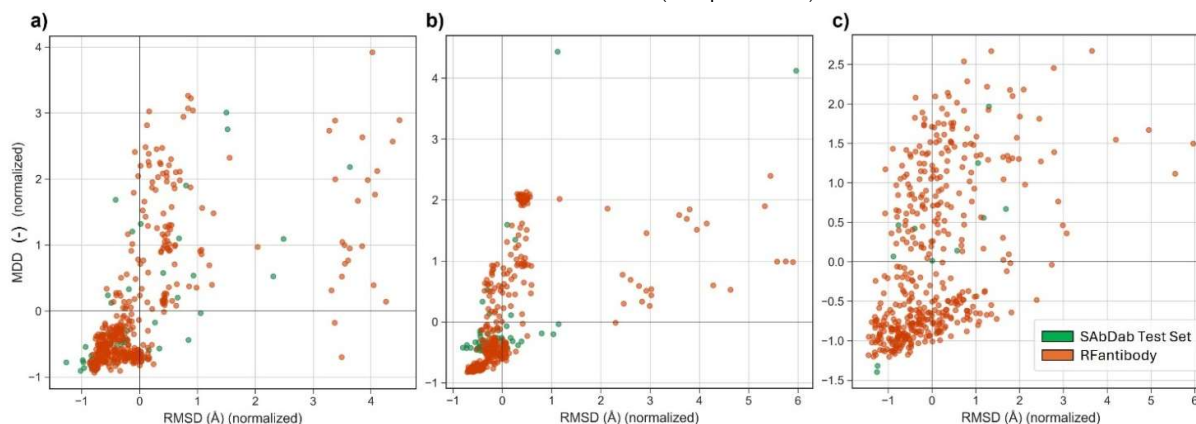


Figure 5. RMSD-MDD Normalised Correlation between values of closest reference match for a) CDR1, b) CDR2 and c) CDR3.

However, some outliers were also evident, showing relatively high RMSD values despite low MDD as seen in Figure 5a. When evaluating the percentage of identical closest matches, we found that RMSD and MDD often identify different matches: for example, in 89.6% of cases for CDR1 in the native test set and 59.7% for RFantibody CDR1 (Table 2)

Table 2. Percentage of same closest RMSD and MDD matches for native and RFantibody generated nanobodies.

	CDR1	CDR2	CDR3
SAbDab Test Set	10.4%	0%	14.3%
RFantibody	40.3%	9%	19.7%

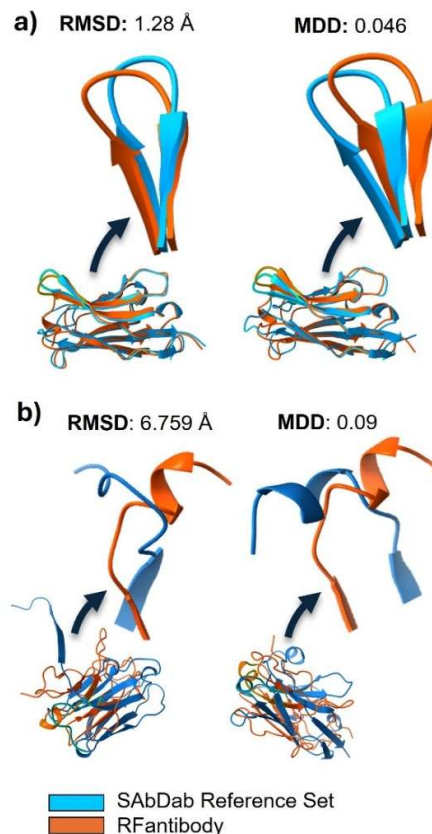


Figure 6. a) Closest RMSD (8emy_G1-116_NA) and MDD (9nk9_A1-115_NA) matches in reference superimposed to nanobody query (8rvr_G1-130_NA_3) for positively correlated points in CDR3 (expected case) and b) Closest RMSD (9g13_E4-127_NA) and MDD (8eye_D1-127_NA) matches in reference superimposed to nanobody query (9g97_B3-127_NA_4) for negatively correlated points in CDR1 (unexpected case).

Even when both MDD and RMSD identify a query as native, with low MDD and RMSD values (0.046 and 1.28 Å respectively) the closest reference matches can differ (Figure 6a). Looking at the expected case, the RMSD-based match aligns closely at the loop's start and end points, diverging mainly at the loop apex. In contrast, the MDD-based match preserves the overall loop conformation but exhibits a noticeable frameshift, explaining its higher RMSD. This likely reflects the positional independence of dihedral angles: small local differences can accumulate, producing large deviations in residue coordinates and thus elevated RMSD.

When investigating outliers—where MDD and RMSD disagree on nativeness (e.g., RMSD = 6.759 Å, MDD = 0.089, as shown on Figure 6b)—show a different pattern. Here, the RMSD-based match is overall dissimilar, while the MDD-based match maintains the same loop pattern but flipped. Because MDD measures position-independent dihedral angles, it is insensitive to CDR chirality, accounting for this phenomenon.

4. Discussion

In this work, we benchmarked the structural nativeness of CDR backbones generated by two diffusion-based antibody design models, RFantibody and the more recent RFD3, against high-quality native structures from SAbDab. Curated reference and test sets ensured evaluation against previously unseen native antibodies, avoiding training leakage, and enabling a fair comparison. Using both positional (RMSD) and angular (MDD) metrics, we assessed conformational similarity across CDRs.

Several trends emerged. First, native test structures consistently clustered nearer to the reference set than RFantibody, validating our approach. Second, RFantibody showed systematic departures from native conformational space, including elevated RMSD and MDD values, Ramachandran outliers, and non-native CDR length combinations. Third, although RFD3 was evaluated only on CDR3, its distributions aligned far more closely with the native test set, in line with expectations for an all-atom generative model.

Correlation analyses further revealed that RMSD and MDD capture distinct aspects of loop geometry: RMSD reflects global spatial arrangement, whereas MDD highlights local angular patterns. Their frequent disagreement on nearest neighbours underscores the need for multi-metric nativeness assessment.

Overall, these results highlight limitations of backbone-only trained diffusion models and using structural information outside of framework and germline context. We establish a foundation for using RMSD-MDD as a practical basis for assessing structural nativeness.

4.1 Native and generated CDR conformations: RMSD/MDD structural assessment

We showed that the RMSD and MDD provide a significant quantitative assessment of how closely each dataset's structures resemble the native reference backbones. For both CDR1 and CDR2, the test set consistently showed lower RMSD and MDD values than RFantibody, indicating that simple structural metrics can distinguish between native and generated backbones. The larger

separation between the test and RFantibody distributions in CDR2 suggests that the generative model struggles particularly with reproducing typical conformations of this loop. CDR2 is known to adopt a much narrower and more tightly clustered set of backbone geometries than CDR1²⁸, reflecting constraints imposed by the germline framework. This likely explains why such low RMSD and MDD values for both the reference and test sets were found for CDR2, as their loops occupy the same narrow region of conformational space. In contrast, despite being trained on real structures that reflect this pattern, RFantibody does not enforce these constraints during generation and therefore samples a broader, less restricted set of geometries. Given the restricted diversity of native CDR2 backbones, deviations from these preferred conformations led to larger RMSD and MDD values in the RFantibody set, while the test set maintained close matches with the reference distribution.

These observations are consistent with broader considerations related to framework-loop compatibility in diffusion-based generative models. In native nanobodies, the CDR1 and CDR2 are encoded entirely within the V gene and the CDR3 at the junction of the V and J genes, providing its diversity⁷. As a result, CDR1 and CDR2 are less variable and more closely related^{7,28}. When generating structures with a fixed framework and user-specified CDR loop lengths, this biological relationship is not considered. This can yield implausible combinations of CDR1 and CDR2 lengths (Supplementary Figure 3.1) or geometries that do not occur in natural repertoires. Adopting a broader or germline-matched set of frameworks or jointly considering CDR1 and CDR2 geometries in a combined structural constraint, may help capture this relationship in future work without requiring explicit sequence information.

The behaviour of CDR3 provided insights into how RMSD and MDD capture different aspects of structural nativeness. For CDR3, both RMSD and MDD separated the datasets less clearly, reflecting the higher intrinsic diversity of this loop. This reduced separation indicates a limitation of structural nativeness metrics when applied to highly diverse loops whose conformational landscape is not well represented in current structural databases.

The closer alignment of the RFD3 distribution with the test set across both metrics indicates that this newer diffusion model may more reliably capture native-like backbone features and diversity within CDR3. This improvement is likely attributable to its all-atom generative architecture. While RFantibody is trained only with amino acid-level information, the RFD3 model is trained on atomic-level information, enabling it to model more native-like, side-chain compatible structure conformations or antigen-interacting loops¹³.

By contrast, the combination of comparable MDD but higher RMSD for RFantibody implies that, although RFantibody can approximate native backbone conformations, it struggles to place CDR3 loops correctly in space relative to the framework, which could affect binding ability.

The partially conflicting RMSD/MDD trends, together with the closer distributions between the native and non-native sets, highlight that these purely structure-based metrics have limited power to distinguish between native

and non-native CDR3 loops. This limitation likely reflects both the conformational diversity of the CDR3 loops, and the relatively small size of available structural datasets compared with sequence datasets, making it difficult to represent a full range of native conformations in our reference set. In contrast to the size of the generated sets, the nanobody test set contained 49 structures with experimentally determined coordinates, reflecting the difficulty in obtaining 3D structure databases, which are limited by high cost and technical demands of crystallisation²⁴. As more structural data becomes available, more statistical power will be available for comparing native datasets to generated ones. Additionally, further studies should examine whether the same RMSD and MDD relationships hold for CDR3 loops longer than those present in the RFD3 dataset.

Based on these findings, RMSD appears to be the preferred metric for distinguishing between native and non-native structures, showing clear separation across all three CDRs. Because RMSD is computed from the optimal superposition of α – carbon coordinates between a query and a reference, it implicitly encodes the global spatial arrangement of CDRs relative to the framework. In contrast, the angular MDD metric, while effective at capturing local conformational features, is less sensitive to global shifts or loop placements that affect paratope geometry. Generative models may therefore produce locally plausible dihedral patterns while failing to account for the integration of a CDR within its framework.

A notable limitation, however, is that RMSD is substantially more computationally expensive due to the required structural superposition (~0.068 s per query–reference pair), compared with MDD (~0.002 s per comparison). To mitigate this bottleneck, RMSD-based structural search accelerators, such as MASTER²⁵, could be incorporated into our scoring pipeline to reduce runtime while preserving the global structural information that appears most predictive of nativeness.

4.2 RMSD and MDD metrics complementarity

Correlation and qualitative analysis of RMSD and MDD closest matches confirmed that, while the two metrics are positively correlated, they capture different aspects of the structural fidelity between generated and native structures. RMSD reflects overall structural resemblance and spatial positioning, whereas MDD better captures local CDR loop patterns. Because MDD is based on dihedral angles and is inherently position-independent, it focuses on local conformational changes. However, it may fail to detect global structural shifts, chirality, or extended loop rearrangements.

Although the metrics offer complementary perspectives on backbone nativeness, RMSD may better predict global epitope engagement from a binding standpoint. The relative positioning and orientation of CDR loops, encoded in RMSD and influenced by variable-domain scaffold orientation, are known determinants of antigen recognition and docking, potentially more so than the local geometric changes highlighted by MDD²⁹. Accurately capturing relative CDR orientation information in a structural nativeness definition is also crucial for assessing developability, as solubility is strongly affected by even slight changes to relative CDR orientation⁶. At the

same time, understanding local CDR loop conformations remains essential: local motifs dictate the exposition of hydrophobic side chains, and large hydrophobic patches in CDRs are directly linked to high aggregation propensity, poor solubility, and high viscosity, all undesirable traits in therapeutic antibodies³⁰. Both metrics therefore provide complementary structural information. It is important to note, however, that neither metric fully accounts for functional binding geometry, highlighting the potential need for an additional binding-aware structural metrics³¹.

To gain finer-grained insight, segment-level MDD (computed over 3–9 residue fragments) can localise recurrent structural motifs and highlight anomalies in generated CDRs relative to solved structures^{7, 32}. Examining average CDR performance across a candidate antibody may provide additional nativeness information. However, assessing functional binding may require additional binding-aware descriptors, such as heavy–light chain torsion angles or epitope contact maps, encoding relative chain orientation and paratope-epitope interactions respectively³³.

This work used an average metric for comparing dihedral distances, providing a single summary score for each CDR loop. While this provides an insight to the overall backbone similarity, it reduces the sensitivity to local conformational difference such as peptide flips⁷. Alternative dihedral-based methods, such as using the maximum phi or psi deviation found within a loop⁷ or using these metrics to create clusters of commonly found conformations are considerations for a future, more developed structural scoring.

Modern machine-learning approaches further could enhance this analysis. Structural descriptors from antibodies can be tokenised and fed into self-supervised learning models, enabling the detection of implicit structural footprints of native CDRs. Transformer architectures, for example, could model inter-CDR relationships and AHo positional dependencies, supporting anomaly detection in generated antibodies. Models such as IgLOO-LM³ demonstrate how data-driven embeddings of antibody structures can capture subtle conformational patterns; similar frameworks could be adapted to identify deviations likely to compromise binding.

5. Conclusion

In this work, we set out to define a structural notion of antibody and nanobody “nativeness” and to use this definition to quantitatively assess the performance of backbone-generating diffusion models. By comparing native structures with RFantibody- and RFD3-generated backbones using two complementary metrics, RMSD and MDD, we evaluated how closely generated CDR loops reproduce experimentally observed conformations.

Across CDR1 and CDR2, native test structures consistently achieved lower RMSD and MDD values than RFantibody, indicating closer agreement with the reference distribution of solved backbones. RFantibody frequently produced loops occupying geometries that were rare or absent in native repertoires, particularly in CDR2, whose canonical conformational constraints were not captured by its framework-agnostic generation strategy. For CDR3, both RMSD and MDD separated the datasets

less clearly, reflecting the higher intrinsic diversity of this loop. Nevertheless, RFD3 backbones were consistently more similar to the native set than those from RFantibody, suggesting that newer diffusion models better capture both global and local structural features of long and diverse CDR3 loops. Taken together, these systematic structural differences between native and non-native backbones provide a clear basis for defining structural nativeness.

Correlation analyses further revealed that RMSD and MDD quantify different aspects of structural nativeness: RMSD is sensitive to global spatial arrangement and framework-loop compatibility, whereas MDD reflects local backbone geometry independent of alignment. The low percentage of shared closest matches between the two metrics confirms that neither alone provides a complete description of structural similarity. Instead, their complementary behaviour supports a combined metric-based definition of nativeness, and highlights the potential of dihedral-based, segment-level analyses to reveal fine-grained conformational deviations not captured by global RMSD.

Several limitations shaped this study. First, the available pool of experimentally solved nanobody structures remains small, constraining statistical power. Second, the generative models did not produce CDR length distributions fully overlapping with native datasets, particularly for long CDR3s, which limited direct comparison. Finally, the use of fixed frameworks in generative pipelines disregards germline constraints linking CDR1 and CDR2 geometry, likely contributing to non-native conformations in RFantibody backbones.

Future work could address these limitations by generating length-matched models with RFD3, expanding the structural reference space as new antibody structures become available, and incorporating framework-loop compatibility directly into backbone generation. Extending dihedral metrics to per-segment or motif-based descriptors, integrating heavy-light chain torsion angles, or encoding nativeness through contact-map embeddings or structural language models represent promising directions for developing richer and more discriminative nativeness scores.

Overall, this study demonstrates that structural nativeness can be quantitatively defined and that RMSD and MDD together form a useful basis for evaluating the quality of generated antibody backbones. The improved performance of RFD3 relative to RFantibody highlights the rapid progress of diffusion models toward generating native-like antibody structures and underscores the importance of rigorous structural metrics for guiding future de novo antibody design.

6. Methods

6.1 Defining native datasets

Structural data for antibodies and nanobodies were obtained from SAbDab, which initially included 4,573 unique paired VH/VL antibody complexes and 1,133 unique VHH nanobody structures. Protein Data Bank (PDB) files were obtained for each structure available, which contained information such as sequences, atomic coordinates, and the chains present.

Entries were filtered to retain only structures with crystallographic resolution below 3.5 Å, complete non-terminal residues, bound complexes, and structures compatible with standard AHO numbering. The remaining structures were split into reference and test sets. The test sets consisted of structures published after December 1, 2024, corresponding to the final training cutoff of RFD3. Additional filters were applied to ensure antibody diversity: test antibodies required a combined CDR chemical distance greater than 50 with at least five CDR mutations relative to any other test structure, and a CDR3 chemical distance greater than 50 with at least three mutations. Antigen diversity was also enforced by limiting any single antigen cluster (e.g. SARS-CoV) to fewer than two representatives within the test set. Structures excluded during these steps were reincorporated into the reference sets.

6.2 Defining generated datasets

Generated antibodies were obtained using RFantibody and RFD3. For each test antigen, epitopes were extracted, and generation tasks were run on predefined framework scaffolds: trastuzumab for RFantibody Fv models, nb156 for RFantibody VHHs, and caplacizumab for RFD3 VHHs, to which the generated CDRs would be grafted.

RFantibody generated CDR backbones across standard length ranges, producing roughly 10 backbones per antigen. Using the same antigens and epitopes, RFD3 structures were provided directly by the Baker Lab and consisted exclusively of VHH backbones. The CDR lengths and generation settings for these RFD3 structures were determined internally, and we did not have control over the specific length ranges used. 50 backbones per antigen were generated using RFD3.

6.3 Restricting length ranges and CDR of matching lengths

Since the test, reference, and generated sets spanned different CDR length ranges (see Figure 2), comparisons were restricted to the length ranges generated by RFantibody, as the primary focus of the diffusion analysis was on this model. Furthermore, to ensure meaningful results and avoid discarding structural information during RMSD and MDD calculations, CDR loops were only compared to loops of the same length⁷.

6.4 Structural comparison methods

6.4.1 RMSD Calculations

To quantitatively assess the structural similarity of a query protein to the reference set, we calculated RMSD of the variable region backbone α -carbons between query and reference structures. For antibodies, the heavy and light chains were superimposed independently, whereas for nanobodies only the heavy chain was aligned, using PDB.Align. RMSD values were computed separately for four regions of the variable domain—CDR1, CDR2, CDR3, and the framework—as defined in Equation 1, where N denotes the number of matching α -carbons in the selected region and i and j the reference and query, respectively.

$$RMSD = \sqrt{\frac{1}{N} \sum_{r=1}^N (x_r^i - x_r^j)^2 + (y_r^i - y_r^j)^2 + (z_r^i - z_r^j)^2} \quad (1)$$

6.4.2 MDD Calculations

Another metric used to quantify structural nativeness was the deviation of backbone angles ϕ and ψ within the CDR regions between query and reference structures. Dihedral angles are a standard way to define backbone geometry and describe the relative orientation of two peptide planes around a central $C\alpha$. The ϕ angle, defined by the planes formed by the atoms $C_{i-1}-N_i-C\alpha_i$ and $N_i-C\alpha_i-C_i$, and ψ is defined by the planes $N_i-C\alpha_i-C_i$ and $N_{i+1}-C\alpha_{i+1}-C_{i+1}$. C_i atoms refer to the acyl carbon⁷. These can be plotted against one another on a Ramachandran plot, which provides visual insights into the secondary structures present such as β -sheets and α -helices²⁷.

An advantage of analysing dihedral angles over RMSD is that each dihedral angle is intrinsically position independent, so no explicit superimposition of query and reference structures is required, only a consistent way to refer to residue positions. In this work, we used the standard AHo numbering scheme.

For two structures, the distance between dihedral angles of the same type, ϕ or ψ , was computed using a directional statistics-based angular distance between angles θ_i and θ_j at the same AHo position³⁴:

$$D(\theta_i, \theta_j) = 2(1 - \cos(\theta_i - \theta_j)) \quad (2)$$

This is based on the squared distance of the chord on a unit circle and accounts for the periodicity of these torsion angles⁷. The per-residue distances, D , were then averaged over the number of residues present in the loop, N , to yield a mean dihedral distance over a given CDR loop, with a maximum value of 4. Only loops of equal length were compared^{7, 34}.

$$\bar{D}(i, j) = \frac{1}{2N} \sum_{r=1}^N D(\phi_r^i, \phi_r^j) + D(\psi_r^i, \psi_r^j) \quad (3)$$

7.X Correlation metrics

To investigate the relationship between the MDD and RMSD metrics, we computed two correlation coefficients: Pearson, r , and Spearman, ρ_s , as defined in Equations 4 and 5, respectively. The Pearson correlation measures the linear relationship between two datasets, while the Spearman correlation assesses a monotonic relationship based on the rank order of the data. Both metrics range from -1 to 1 , where values near 1 indicate a strong positive correlation (linear for Pearson, monotonic for Spearman), and values near -1 indicate a strong negative correlation.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

Where x_i is the value of variable x for observation i and y_i the value of y for observation i , $\bar{x}$ and $\bar{y}$ the averages of both x and y variables and n is the number of data points.

$$\rho_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (5)$$

Where d_i is the difference in ranks between each paired information and n the number of paired observations.

References

- [1] Sonal, Rana, R., Dahiya, N., & Kaur, P. (2024). The Role of Antibodies in Therapeutics: A Biotechnological Perspective. *Advances in BioResearch Special Issue 1*, 141–144.
https://www.researchgate.net/publication/393550880_The_Role_of_Antibodies_in_Therapeutics_A_Biotechnological_Perspective
- [2] Precedence Research. (2025). Antibody Therapy Market Size and Forecast 2025 to 2034. <https://www.precedenceresearch.com/antibody-therapy-market>
- [3] Fang, A., et al. (2025). Tokenizing loops of antibodies. *arXiv*.
<https://doi.org/10.48550/arXiv.2509.08707>
- [4] Kim, J., et al. (2023). Computational and artificial intelligence-based methods for antibody development. *Trends in Pharmacological Sciences*, 44(3), 175–189.
<https://doi.org/10.1016/j.tips.2022.12.005>
- [5] Sormanni, P., et al. (2018). Third generation antibody discovery methods in silico rational design. *Chemical Society Reviews*, 47(24), 9137–9157.
<https://pubs.rsc.org/en/content/articlelanding/2018/cs/c8cs00523k>
- [6] Wolf Pérez, et al. (2022). Assessment of Therapeutic Antibody Developability by Combinations of In Vitro and In Silico Methods. <https://doi.org/10.17863/CAM.75474>
- [7] Kelow, S., et al. (2022). A penultimate classification of canonical antibody CDR conformations. Preprint.
<https://doi.org/10.1101/2022.10.12.511988>
- [8] Mitchell, L. S., & Colwell, L. J. (2018). Comparative analysis of nanobody sequence and structure data. *Proteins*, 86(7), 697–706.
<https://doi.org/10.1002/prot.25497>
- [9] Bashour, H., et al. (2024). Biophysical cartography of antibody landscapes. *Communications Biology*, 7(1), 922.
<https://doi.org/10.1038/s42003-024-06561-3>
- [10] Ramon, A., et al. (2024). Assessing antibody and nanobody nativeness with AbNatiV. *Nature Machine Intelligence*, 6(1), 74–91.
<https://doi.org/10.1038/s42256-023-00778-3>
- [11] Kim, J. & Hong, H. (2012). Humanization by CDR Grafting and SDR Grafting. *Antibody Engineering*, 907, 237–245.
https://doi.org/10.1007/978-1-61779-974-7_13
- [12] Bennett, N. R., et al. (2025). De novo antibody design with RFdiffusion. *Nature*.
<https://doi.org/10.1038/s41586-025-09721-5>
- [13] Kenneth, J., et al. (2025). RFdiffusion3. *BioRxiv*.
<https://doi.org/10.1101/2025.09.18.676967>
- [14] Mille-Fragoso, L., & John, N. (2025). Germinal: epitope-targeted de novo antibodies. Preprint.
<https://doi.org/10.1101/2025.09.19.677421>
- [15] Stark, H., Faltings, et al. (2025). BoltzGen: Toward universal binder design. Preprint.
<https://hannes-stark.com/assets/boltzgen.pdf>
- [16] Høie, M., & Hummer, G. (2025). AntiFold: improved structure-based antibody design. *Bioinformatics Advances*, 5(1). <https://doi.org/10.1093/bioadv/vbae202>
- [17] Ruffolo, J. A., et al. (2023). Fast, accurate antibody structure prediction. *Nature Communications*, 14, 2389.
<https://doi.org/10.1038/s41467-023-38063-x>
- [18] Rollins, Z. et al. (2024). AbMelt: Learning antibody thermostability. *Biophysical Journal*, 123(17), 2921–2933.
<https://doi.org/10.1016/j.bpj.2024.06.003>
- [19] Ramon, A., et al. (2025). NanoMelt: Nanobody thermostability prediction. *MAbs*.
<https://doi.org/10.1080/19420862.2024.2442750>
- [20] Marks, C., et al. (2021). Humanization using machine learning. *Bioinformatics*, 37(22), 4041–4047.
<https://doi.org/10.1093/bioinformatics/btab434>
- [21] Kayser, V., et al. (2011). Conformational stability of monoclonal antibodies. *MAbs*, 3(4), 408–411.
<https://doi.org/10.4161/mabs.3.4.15677>
- [22] SAbDab: The Structural Antibody Database. (2022). *Ox.ac.uk*. <https://opig.stats.ox.ac.uk/webapps/SAbDab-sabpred/SAbDab/about>
- [23] Olsen, T. H., et al. (2021). Observed Antibody Space. *Protein Science*, 31(1), 141–146.
<https://doi.org/10.1002/pro.4205>
- [24] Grey, J. L., & Thompson, D. H. (2010). Protein crystallization challenges. *Expert Opin Drug Discov*, 5(11), 1039–1045.
<https://doi.org/10.1517/17460441.2010.515583>
- [25] Zhou, J., & Grigoryan, G. (2015). Rapid search for tertiary fragments. *Protein Science*.
<https://doi.org/10.1002/pro.2610>
- [26] Shao, C., et al. (2022). Assessing PDB macromolecular crystal structure precision using real-space measures. *Structure*, 30(12), 1751–1762.e3.
<https://doi.org/10.1016/j.str.2022.08.004>
- [27] Brzozowski, K. (n.d.) Ramachandran Plot as a Tool for Peptide and Protein Structures' Quality Determination. *PeptideWeb*. Available at: <https://www.peptideweb.com/ramachandran-plot-as-a-tool-for-peptide-and-protein-structures-quality-determination>
- [28] Teplyakov, A., et al. (2016). Structural diversity in germline antibody libraries. *MAbs*, 8(6), 1045–1063.
<https://doi.org/10.1080/19420862.2016.1190060>
- [29] Dunbar, J., et al. (2014). Variable domain orientations in antigen receptors. *PLoS Computational Biology*, 10(9), e1003852. <https://doi.org/10.1371/journal.pcbi.1003852>
- [30] Zhang, W., et al. (2022). Developability assessment in early-stage discovery. *Antibody Therapeutics*.
<https://doi.org/10.1093/abt/tbac029>
- [31] Sela-Culang, I., et al. (2013). Structural basis of antibody–antigen recognition. *Frontiers in Immunology*, 4.
<https://doi.org/10.3389/fimmu.2013.00302>
- [32] Sawada, Y., & Honda, S. (2006). Structural diversity of protein segments. *Biophysical Journal*, 91(4), 1213–1223.
<https://doi.org/10.1529/biophysj.105.076661>
- [33] Dunbar, J. B., et al. (2013). ABangle. *Protein Engineering Design & Selection*, 26(10), 611–620.
<https://doi.org/10.1093/protein/gzt020>
- [34] North, B., et al. (2011). Clustering of CDR loop conformations. *Journal of Molecular Biology*, 406(2), 228–256. <https://doi.org/10.1016/j.jmb.2010.10.030>

Atmospheric Plasma-Induced Wrinkling as a Scalable Surface Patterning Technique

Robert Middleton^a, Ayushman Kashyap^a

Surface patterning at the micro- and nanoscale by controlled wrinkling of elastomeric materials underpins diverse technological applications. However, existing methods remain limited by high operational costs and infrastructure demands. Atmospheric plasma has emerged as a cost-effective alternative to vacuum plasma and ultraviolet-ozone oxidation, but its viability for producing comparable wrinkle characteristics remains unestablished. We use a piezoelectric direct discharge device to investigate atmospheric plasma-induced wrinkling of polydimethylsiloxane. The effects of plasma dosage and elastomer strain or wrinkle features were investigated while introducing motion-assisted treatment strategies to enhance spatial uniformity. Atmospheric plasma created wrinkle wavelengths and amplitudes in range of those produced by vacuum plasmas, with calculated film thicknesses between 2.26 nm and 34.87 nm. These are consistent with literature values, confirming the frontal vitrification oxidation mechanism developed for vacuum plasma, applies to atmospheric plasma as well. An 80% improvement in wrinkle uniformity within a sample was observed from simple translational motion to fine translational motion at 0.05 mm/s, while rotational motion gave an 85% improvement in wavelength uniformity within a sample relative to the simple translation motion. The results show that atmospheric plasma is a practical and scalable surface patterning technique viable for industry, which offers reduced capital expenditure, operational complexity, and is able to treat complex geometries without vacuum infrastructure. However, further optimisation of translation speed and parallel processing configurations are required to make it a compelling case for complete industrial deployment.

Keywords: wrinkling, PDMS, atmospheric plasma, translational and rotational motion, scalable fabrication

1 Introduction

A diverse array of technological applications have been unlocked from surface patterning at the micro- and nanoscale: from controlled spreading of biological cells¹ and antibacterial action^{2,3} to antifouling coatings⁴, tuneable wetting⁵, structural colour⁶ and drag reduction⁷. To fabricate these functional topographies, surface wrinkling has emerged as an attractive route amongst various non-lithographic approaches due to its low cost⁸, simplicity⁹, and capacity to generate highly regular periodic patterns over large areas without the need for specialised clean room facilities¹⁰. Through in-plane mechanical compression¹¹, prescribed wavelengths and amplitudes can be achieved due to the mechanical instability¹² of bilayer structures comprising a thin, stiff film bonded to a compliant elastomeric substrate¹³. The favourable combination of physical-chemical properties, including low elastic modulus¹⁴ (≈ 1.6 - 2.7 MPa), optical transparency^{15,16}, large elongation at break¹⁴ ($\approx 200\%$), near incompressibility^{17,18} (Poisson's ratio $\nu \approx 0.5$), and readily tuneable surface chemistry¹⁸, has directed considerable attention to the wrinkling of polydimethylsiloxane (PDMS).

The surface of PDMS can be oxidised through several methods, including exposure to ultraviolet-ozone (UV-O) and plasma treatment¹⁹, both which create a dense, silica-like glassy skin by cleaving Si-CH₃ bonds and progressively replacing them with hydroxyl groups and oxide links²⁰, to produce a requisite stiff surface layer. UVO oxidation yields larger wrinkling wavelengths of tens to hundreds of micrometers by converting PDMS with a non-uniform cross sectional profile (50-100nm). In vacuum plasma oxidation, ionisation of gas occurs through an oscillating electric field under reduced pressure²¹, to generate thin glassy layers ($\approx$

10nm) that yield wrinkles with wavelengths ranging from ≈ 100 nm to tens of micrometers upon mechanical compression. The glassy skin, formed by a frontal vitrification mechanism²², can be characterised by distinct induction, formation, and propagation stages (as shown in *Figure 1a*), and the film thickness (h_f) has been shown to increase logarithmically with plasma dose (product of power and exposure time)²². The industry scalability of these techniques remains limited by inherent constraints despite extensive investigation of vacuum and UVO-based approaches, with UVO treatment having limited penetration depth²³, and vacuum systems requiring pump-down cycles²⁴ and controlled environment chambers, incurring significant energy consumption, operational complexity, and capital expenditure²⁵. Therefore, atmospheric plasma treatment, which offers potential for inline integration into automated production lines²⁶, reduced operational costs²⁷, and the ability to treat complex or large geometries²⁸ without vacuum infrastructure, is a promising alternative to the traditional methods discussed above.

In this study, atmospheric plasma-induced wrinkling of PDMS is studied using a piezoelectric direct discharge (PDD) plasma device²⁹ (Piezobrush PZ3) operated at ambient pressure, to determine its scalability as a surface patterning technique. We first assess whether atmospheric plasma oxidation yields wrinkle wavelengths and amplitudes consistent with those produced using vacuum plasma and UV-ozone. We then examine the role of treatment dose and applied strain on wrinkle characteristics, followed by the introduction of controlled rotational and translational motion to improve spatial uniformity and evaluate the feasibility of scaling the process. Together, these results aim to establish the viability of atmospheric plasma as a practical, large-area alternative for generating mechanically tunable wrinkled surfaces.

^a Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK

* Corresponding author. Email: robert.middleton22@imperial.ac.uk

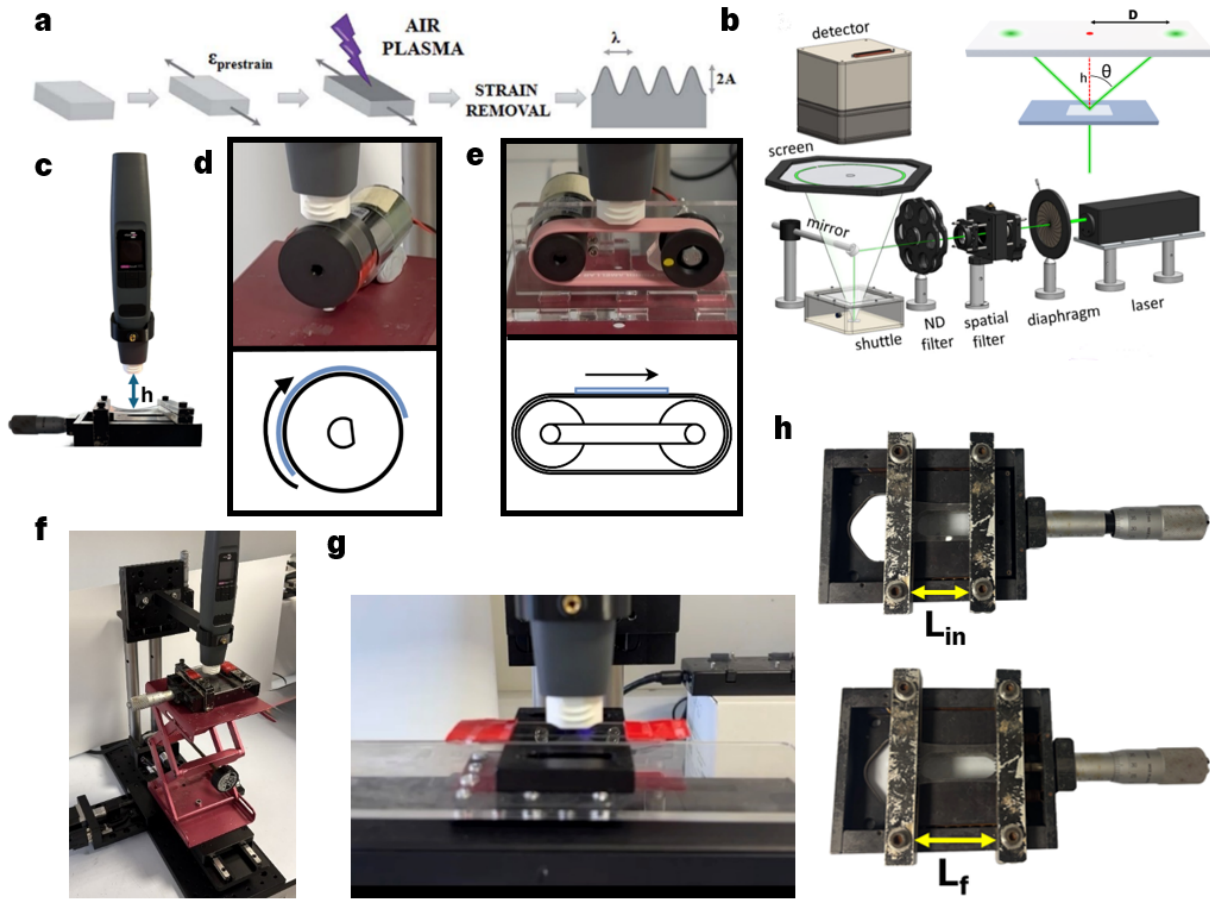


Figure 1 **a**, Formation of wrinkles on SiO_x -PDMS bilayers by plasma oxidation: PDMS specimen is pre-strained uniaxially by ϵ_{pre} , then exposed to atmospheric plasma for varying times, and then allowed to relax forming sinusoidal wrinkling patterns with prescribed λ and A from which the glassy skin thickness h_f is inferred²² **b**, A depiction of the small angle laser scattering setup¹⁹ **c**, A depiction of the piezobrush module-to-PDMS surface distance **d**, A representation of the rotational motion set-up **e**, A representation of the roll-to-roll setup for continuous processing. **f**, A depiction of the piezobrush, held by our 3-D printed holder, with the height adjusting gauge allowing us to accurately quantify the module-to-PDMS surface distance **g**, Our fine translations setup with PDMS stretched on a glass backing **h**, Images of the uniaxial strain stage used: pre-strain and post-strain

2 Methodology

2.1 Substrate Fabrication and Pre-Strain Procedure

PDMS substrates were prepared using a ratio of base:curing agent of 10:1, giving a Young's modulus that was practical for handling and experimentation. The mixture was cured according to manufacturers recommendations and cut into samples of size $\approx 1.5 \text{ cm} \times 5 \text{ cm}$ before being clamped into a strain stage. The stage allowed for a controlled uniaxial pre-strain to a calculable percentage by measuring the initial and final length of the PDMS as per Equation 1:

$$\epsilon(\%) = \frac{L_{\text{final}} - L_{\text{initial}}}{L_{\text{initial}}} \times 100\% \quad (1)$$

To verify that atmospheric plasma can induce wrinkling, initial trials were conducted with a fixed arbitrary strain of 20% $\approx 4 \text{ mm}$ plasma module to PDMS distance and 5 minutes exposure time. Keeping plasma percentage power at 100% with a known intensity of 8 W and a measurable exposure time, plasma

dose as a unit of energy can be calculated:

$$\text{Plasma dose (J)} = 8 \text{ W} \times \text{Exposure time (s)} \quad (2)$$

A simple visual inspection is sufficient to confirm the presence of wrinkling, with the nanoscale topography producing observable optical diffraction patterns visible to the naked eye.

2.2 Plasma Dose Investigation

To investigate the effect of plasma dose on the wrinkle wavelength and amplitude, a simple translational motion was introduced to the setup to provide a more even dose across the surface of the sample than could be achieved with a stationary setup. While this motion was not optimised for uniformity of wrinkle characteristics, it provided a reproducible central patch on each sample where measurements using Atomic Force Microscopy (AFM) could be taken. In addition to the programmable moving XY stage, an adjustable Z-stage fitted with a micrometer was introduced to quantify the plasma module to sample surface distance, h (mm).

Wrinkle wavelengths and amplitudes were measured after

plasma exposures of 1, 10, 20, 30 and 60 minutes, following removal of the applied strain. For each series of exposure times, the module to sample height was kept constant; different series were conducted at heights of $h = 2, 4, 6, 8$ mm. The selection of these heights is justified by the manufacturer's guidance, which specifies a suitable operating range of 1 – 10 mm. Exposure times from 1 to 60 minutes ensured that the complete temporal evolution of wrinkle formation was captured, allowing for a representative curve as discussed in Section 3.3.

2.3 Elastomer Strain Investigation

Wrinkle amplitude is most affected by the percentage of strain applied to the PDMS, and the effect on wavelength is less pronounced³⁰. This investigation focuses only on the relationship between PDMS pre-strain and amplitude. Pre-strain provides a separate handle to control amplitude, allowing tunable wrinkle features, which is particularly relevant for industrial applications³¹.

In this study, all samples were investigated under constant plasma treatment conditions: a fixed module-to-PDMS height of 4 mm and an exposure dose of 9600 J (corresponding to a 20-minute treatment). Pre-strains of 5%, 10%, 20%, and 30% were selected to ensure that all samples exceeded the critical strain required for wrinkling. The onset of wrinkling in a stiff-soft bilayer can be estimated using the classical critical strain expression:

$$\epsilon_c = \frac{1}{4} \left(\frac{3E_s}{E_f} \right)^{2/3} \quad (3)$$

where E_f and E_s are the moduli of the oxidised silica-like skin layer and the PDMS substrate, respectively. For plasma-oxidised PDMS, the large modulus contrast ($E_f/E_s \sim 10^5$ – 10^6) results in a very small critical strain, typically on the order of 0.1%–1%, as reported in the literature³².

This indicates that even the lowest strain used in this study (5%) lies comfortably above the wrinkling threshold, ensuring the formation of measurable wrinkle amplitudes and wavelengths. The upper bound of 30% was dictated by the mechanical limitations of the strain stage and the tensile robustness of the PDMS films, which began to tear at strains exceeding $\approx 30\%$.

2.4 Analytical Techniques of Measuring Wrinkle Characteristics

2.4.1 Atomic Force Microscopy

Atomic Force Microscopy (AFM) was employed to quantify the wrinkle topography generated on the plasma-treated PDMS samples. A central section of each sample, approximately 8 mm $\times$ 8 mm, was cut for AFM analysis, as this region received the most uniform plasma dose and could be mounted securely onto the AFM sample puck. Within this region, scans of area size 50 μ m $\times$ 50 μ m were acquired. This scan size was selected to capture multiple wrinkle periods across all wavelengths observed in the study, ensuring that wavelength and amplitude measurements were statistically representative rather than based on a single wrinkle cycle. For each condition, multiple scans were collected to account

for spatial variability, and the resulting height profiles were processed using standard line-profile analysis. These AFM measurements provided the primary quantitative dataset used to compare wrinkle characteristics across plasma dose, pre-strain level, and motion-assisted treatment conditions.

2.4.2 Laser Diffraction

Laser diffraction, using a small-angle light scattering (SALS) system, was employed to rapidly assess wrinkle wavelengths and spatial uniformity across the full PDMS sample surface. The SALS apparatus, consisted of a diode-pumped green laser, reflected on to a mirror to direct the beam normal to the PDMS surface. When incident on the sample, the sinusoidally wrinkled surface acts as a diffraction grating, producing a characteristic series of bright diffraction maxima on a screen. The vertical distance between the sample and the screen was altered to show only the first-order diffraction pattern, with the central maxima blocked out by a beam stop. The spacing of these peaks is directly related to the wrinkle wavelength through the grating equation:

$$n\lambda = d \sin \theta, \quad (4)$$

where n is the diffraction order, λ is the laser wavelength, d is the wrinkle wavelength, and θ is the scattering angle of the corresponding diffraction peak. Measuring θ therefore provides a rapid, non-contact means of determining wrinkle periodicity and assessing its uniformity across the sample.

Unlike AFM, which provides high-resolution local topography, the SALS technique enabled rapid evaluation of wrinkle wavelength variation across the entire sample. Diffraction patterns were collected across the entire sample to visualise variations in wrinkle uniformity, as changes in fringe spacing and intensity directly indicated local variations in wrinkle wavelength and ordering. This allowed non-uniform plasma exposure to be detected immediately.

To systematically characterise spatial variability, nine measurement points were selected across each sample in a 3 $\times$ 3 grid arrangement, evenly distributed across the sample surface, ensuring no points were too close to the edge. This sampling strategy provided a representative map of surface periodicity and enabled identification of gradients or asymmetries in the wrinkle pattern. It proved particularly useful in our investigation of improving wrinkle uniformity, as it allowed immediate visualisation of variations in wavelengths across the sample.

2.4.3 Optical Microscopy

Optical microscopy was used to visualise and assess the formation of surface cracks on plasma-treated PDMS samples following mechanical strain release. Samples were examined using the Olympus BX41LED optical microscope equipped with reflected-light illumination. Magnifications from 5x to 50x were used to produce visuals of crack morphology across samples. These magnifications helped visualise both localised crack initiation sites, and their propagation across the sample surface.

Images were acquired and analysed using the Pylon Viewer software, which utilised the microscope's integrated camera system to provide visuals of sample surfaces. This real-time

image capture allowed systematic documentation of surface damage across a representative for each motion-assisted treatment method (rotational, roll-to-roll, and fine translational approaches). Images were acquired from at least 3 non-overlapping regions to find the most clear image which was then saved.

The primary purpose of optical microscopy was qualitative assessment of crack density as a function of motion-assisted treatment method. The roll-to-roll method showed widespread cracks disrupting the periodic wrinkle structure, as documented in section 3.5.2. Conversely, samples treated using the rotational method showed a lower density of surface cracks. To identify the mechanical failure mechanisms underlying the different motion-assisted treatment methods, the optical images proved invaluable. Particularly, in the case of the roll-to-roll configuration which involved strain cycling effects. A comprehensive assessment of desired surface topography and undesired crack formation across different experimental conditions was achieved by coupling the qualitative information from the optical microscopy with the quantitative information from the atomic force microscopy (AFM) and small-angle light scattering (SALS).

2.5 Improving Uniformity

2.5.1 Rotational Motion

Rotational motion was introduced as a means to improve the spatial uniformity of the plasma-induced wrinkles. Previous experiments using purely translational motion exhibited significant variation in wrinkle wavelength and amplitude across the sample, limiting both reproducibility and scalability. Incorporating rotational motion addresses these limitations and is particularly relevant for the design of continuous, large-area processing techniques.

In our setup, the rotational motion was implemented using a motor fitted with a custom 3D-printed cylinder of radius $R = 19.9\text{mm}$. The pre-strain was applied by wrapping the PDMS around the cylinder and securing both ends. Taking the initial sample length and using the arc length of the cylinder, the pre-strain could be calculated. Increasing the voltage on the power supply increased the revolutions per minute (rpm) of the cylinder. This approach allowed the plasma exposure to be effectively averaged over the sample surface, enhancing the uniformity of the resulting wrinkle patterns while providing a scalable method suitable for larger-area processing.

2.5.2 Roll-to-roll Motion: Continuous Processing

To explore the potential for scalable surface patterning, a roll-to-roll (R2R) continuous processing setup was developed. This system was designed to enable plasma treatment over extended sample lengths in a controlled and repeatable manner. The apparatus was custom-fabricated using a combination of a laser-cut base structure, 3D-printed rollers, ball bearings and a rubber wrist band to form a compact mini-conveyor belt, capable of transporting pre-strained PDMS samples through the plasma exposure zone at a controlled yet variable speed.

In operation, the pre-strained PDMS samples were mounted on the moving belt, which guided the samples past the atmospheric

plasma source. The pre-strained PDMS samples were mounted on the moving belt to ideally maintain tension to preserve the pre-strain. This motion was intended to allow more uniform interaction with the plasma over the entire surface while demonstrating a scalable approach for larger-area processing. Implementation challenges related to strain management around the rollers are discussed in the Results section.

2.5.3 Fine Translation Motion

To achieve precise control over plasma exposure and wrinkle uniformity, a fine translational motion setup was employed using a high-precision motorised stage with a minimum speed of 70nm/s . Pre-strained PDMS samples, were mounted with their underside adhered to a glass slide to maintain constant strain during treatment. The samples were then translated beneath the atmospheric plasma source, moving at a controlled speed of approximately 0.19mm/s . This motion allowed each region of the sample to receive a consistent plasma dose equivalent, while the fine, continuous movement helped to minimise spatial variations in wrinkle formation across the surface. The setup thus provided a controlled, reproducible means to investigate the influence of motion-assisted exposure on wrinkle uniformity and pattern characteristics.

3 Results and Discussion

3.1 Oxidised PDMS response under stationary conditions

The initial experiments involved a stationary setup to verify that atmospheric plasma was capable of producing well-defined wrinkles with wavelengths and amplitudes comparable to those of other surface oxidation techniques. Figure 2.a shows the characteristic iridescent colouration associated with diffraction, confirming the presence of surface wrinkling. Interestingly, the Figure exhibits a distinct bilobed profile, similar to that of a butterfly shape, with an unexposed central region. This pattern is the intrinsic spatial distribution of the atmospheric plasma module's exposure zone, which produces a non-uniform exposure field. This non-uniform exposure was the key driver for the need to introduce motion to improve uniformity for scalable applications.

3.2 Biaxial strain

The primary focus of the research is on one-dimensional surface topologies, as produced by a strain in a single direction. However, for atmospheric plasma to be a viable alternative to other surface patterning techniques, it must have the potential to produce more complex surface topologies induced by biaxial or multiaxial pre-strain. Figure 2.b shows the resulting surface pattern following plasma exposure (exposure time $t = 20\text{min}$, $h = 4\text{mm}$) when a 20% pre-strain is applied in both directions. The zig-zag patterning is achieved as expected, showing that atmospheric plasma is able to produce more complex surface patterning like vacuum plasma and UV-ozone oxidation techniques achieve.

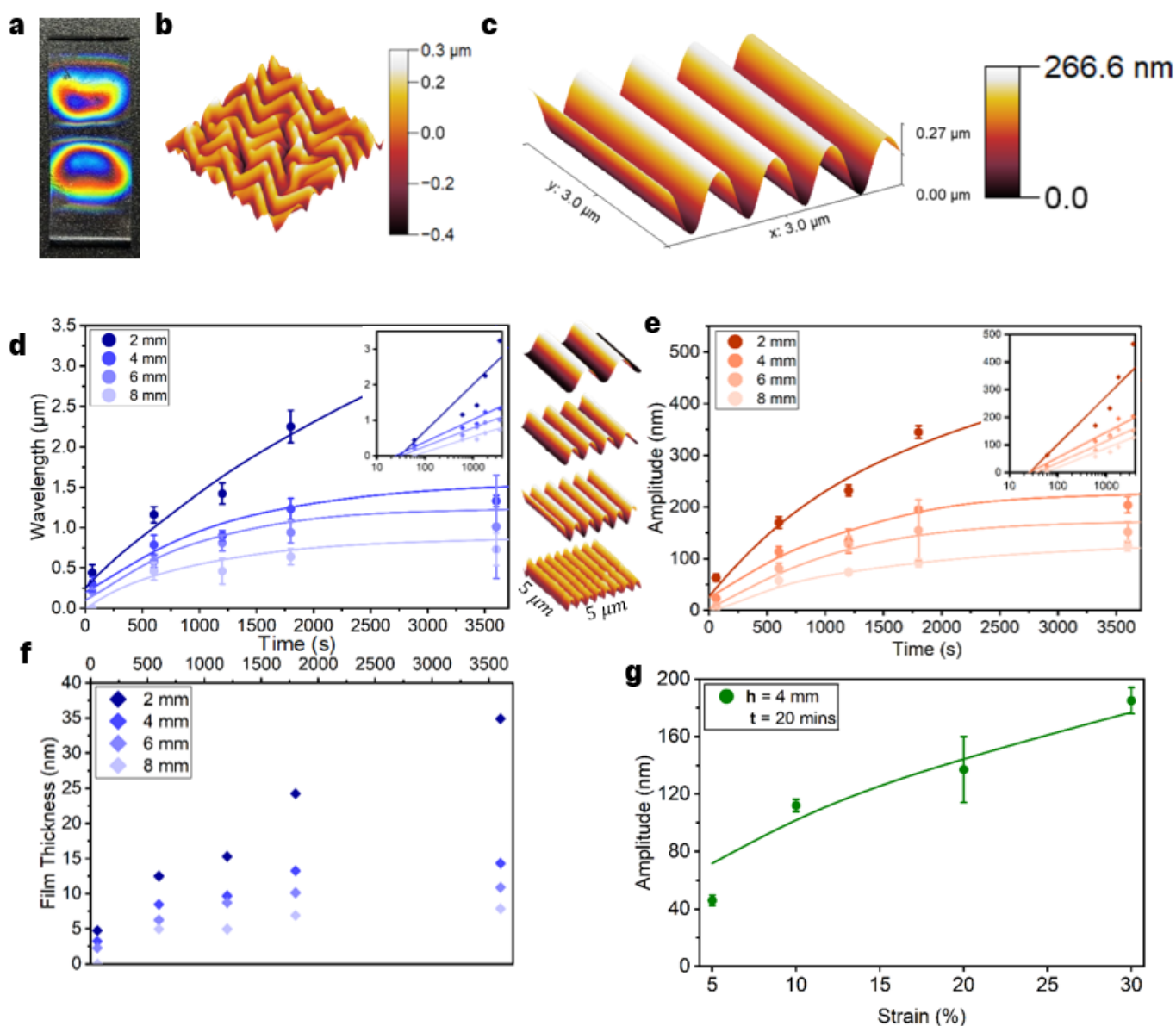


Figure 2 **a**, The sample prepared using the piezobrush stationary with respect to the PDMS surface, depicting a butter-fly shape showing non-uniformities in wrinkling across the sample **b**, A depiction of biaxial strain and its characteristic "zig-zag" shaped wrinkles, showing transferability from vacuum plasma **c**, An example of wrinkle topology obtained using atomic force microscopy, with clear defined peaks and ridges for wavelength and amplitude measurements. **d**, Graph showing wavelength λ plotted against exposure time t , depicting logarithmic shape verified by subplot which graphs wavelength λ against $\ln(t)$, also confirming the established bilayer mechanics relations established for vacuum plasma, and showing its transferability to atmospheric plasma. **e**, Graph showing amplitude A plotted against exposure time t . The logarithmic relationship is verified using the sub-plot which plots amplitude A against $\ln(t)$. The AFM images adjoining the graphs show the wrinkles formed for each module-PDMS surface distance (h) for $t = 1800$ s exposure time. Wrinkles get smaller from $h = 2$ mm to $h = 8$ mm **f** Graph showing film thickness values h_f plotted against exposure time. h_f is calculated from the wavelengths above using the established bilayer mechanics equations, representing a logarithmic shape, verifying the applicability of these equations to atmospheric plasma **g**, Graph plotting amplitude A against PDMS pre-strain ϵ_{pre} , showing a square-root relationship as per the established bilayer mechanics equations.

3.3 Dependence of Wrinkle Characteristics with Atmospheric Plasma Dosage

3.3.1 Wavelength vs Exposure Time

From Figure 2.d, the relationship between wrinkle λ and exposure time t is consistent with the established frontal vitrification mechanism reported for PDMS oxidation by vacuum plasma²². As observed previously in vacuum plasma systems, the following

relationship $h_f \propto \ln(t)$ is observed, indicating that the glassy film thickness h_f increases logarithmically with continued plasma exposure.

The most pronounced wavelength evolution was observed at the shortest PDMS-to-module distance of $h = 2$ mm, ranging from $0.44 \mu\text{m}$ at $t = 60$ s to $3.2 \mu\text{m}$ at $t = 3600$ s. At reduced distances, the higher plasma intensity incident on the PDMS surface accel-

erates the oxidation kinetics and promotes thicker glassy film formation, leading to a substantial increase in wavelengths. This logarithmic trend is characteristic of the diffusion-limited frontal process, where the polymer network is progressively penetrated by ionised species and oxygen radicals, forming Si–O–Si crosslinks and cleaving the Si–CH₃ bonds²². The wavelength evolution becomes progressively less pronounced as the distance increases to $h = 4$ mm, 6 mm, and 8 mm, with smaller final wavelengths achieved at $t = 3600$ s (1.33 μm , 1.01 μm , and 0.73 μm respectively). This indicates that the atmospheric plasma intensity decays with increasing PDMS surface-to-piezobrush module distance, resulting in thinner glassy films and reduced oxidation efficiency.

The established bilayer mechanics relationship for wrinkling in the low-to-moderate deformation regime can be used to interpret the wavelength data in terms of glassy film thickness³³ h_f as shown in Equation 5:

$$\lambda = \frac{2\pi h_f}{(1 + \varepsilon_{pre})(1 + \xi)^{1/3}} \left(\frac{\bar{E}_f}{3\bar{E}_s} \right)^{1/3} \quad (5)$$

where ε_{pre} is the applied pre-strain (20% in this case), $\xi = \frac{5\varepsilon(1+\varepsilon)}{32}$, $\bar{E}$ is the plain strain modulus³³ given by $\bar{E} = \frac{E}{1-\nu^2}$ and the subscripts f and s denote the film and substrate respectively. For the PDMS Young's modulus $E_s = 1.8$ MPa and Poisson's ratio¹⁸ of $\nu_s \approx 0.5$, $\bar{E}_s = 2.3$ MPa, and $\bar{E}_s$ is taken to be 40 GPa, a value consistent with literature for the bilayer low deformation model³². The calculated glassy film thicknesses were calculated to be in the range $h_f \approx 0 - 35\text{nm}$ (0 being for no wrinkle formation at $h = 8$ mm with 60 s exposure time, as shown in Figure 2.d. This is consistent with the typical film thicknesses ($h_f \approx 10 - 50\text{nm}$) reported for PDMS oxidation by vacuum plasma³⁴, confirming that atmospheric plasma treatment produces comparable oxidation depths despite the different exposure geometry.

The frontal vitrification model²² developed for vacuum plasma systems quantitatively describes the logarithmic kinetics in Figure 2.f by the Equation 6:

$$h_f = \frac{\ln(D)}{\mu} - \frac{1}{\mu} \ln \left(\frac{1}{KI_0} \ln \left(\frac{1}{1 - \phi_c} \right) \right) \quad (6)$$

where D is the plasma dose ($D = 8W \times t$), μ is an averaged attenuation coefficient (weighed by composition), I_0 is the incident plasma intensity, and ϕ_c is a critical conversion value required for glass formation (analogous to percolation threshold). Applying this logarithmic relationship to atmospheric plasma data suggests that the fundamental oxidation mechanism of progressive penetration of reactive oxygen species into the PDMS network is analogous to vacuum systems. Spatial control over oxidation depth by adjusting sample position below the module introduces an additional processing parameter otherwise unavailable in chamber-based systems: $I_0(h)$.

3.3.2 Amplitude vs Exposure Time

As seen in Figure 2e, the amplitude of wrinkles A exhibit a logarithmic increase with continued plasma exposure, analogous to the wavelength evolution, and consistent with the frontal

vitrification mechanism underlying PDMS oxidation wherein glassy film thickness h_f increases logarithmically with plasma dose²². The sharp increase of amplitude from 63 nm at an exposure time of $t = 60$ s to 465 nm at $t = 3600$ s at the shortest plasma source-to-sample distance of $h = 2$ mm, reflects the accelerated oxidation kinetics at reduce standoff distances where incident plasma intensity is the highest. At $h = 4$ mm, the amplitude reaches 204 nm at $t = 3600$ s, while at $h = 6$ mm, the amplitude is 152 nm, mirroring the wavelength behaviour observed in Section 3.3.1, indicating that both wrinkle characteristics are governed by resulting glassy film thickness and the spatial intensity distribution of the atmospheric plasma jet.

As shown in Equation 7, the established bilayer mechanics relationship for wrinkling in the low-moderate deformation regime can be used to predict amplitude values if glassy film thickness³³ h_f values are known, and vice versa.

$$A \approx \frac{h_f \sqrt{\varepsilon_{pre}/\varepsilon_c - 1}}{\sqrt{1 + \varepsilon_{pre}(1 + \xi)^{1/3}}} \quad (7)$$

where ε_{pre} is the applied pre-strain (20% in this case), $\xi = \frac{5\varepsilon(1+\varepsilon)}{32}$, and ε_c is the critical strain for wrinkling. As the wrinkle amplitude depends directly on film thickness according to Equation 7, the amplitude is expected to exhibit similar logarithmic kinetics as film thickness as expressed in Equation 6. The logarithmic relationship is verified by the sub-plot in Figure 2.e which plots amplitude against $\ln(t)$ yielding an approximate linear relationship for all distances tested.

Through combined control of module-to-PDMS surface distance and exposure time, wrinkle amplitude can be tuned across a wide range (from tens to hundreds of nanometers). This is particularly valuable for applications requiring prescribed amplitude specifications, enabling optimisation of amplitude through dose adjustment.

3.4 Variation of Wrinkle Amplitude with Elastomer Strain

The wrinkle amplitude was observed to increase with increasing pre-strain, as shown in Figure 2.g. The measured amplitudes ranged from approximately 46 nm at 5% strain to 168 nm at 30% strain, indicating a clear dependence of wrinkle amplitude on the applied substrate deformation as seen in Equation 7.

As the applied strain increases well beyond the critical strain ($\varepsilon \gg \varepsilon_c$), the amplitude grows approximately with the square root of the strain ratio, consistent with theoretical expectations.

The data points exhibit good agreement with this scaling, confirming that pre-strain provides an effective and tunable control of wrinkle amplitude. Some scatter in the measurements is observed, particularly at intermediate strain levels, which may reflect local variations in plasma-induced skin thickness or minor spatial heterogeneities in the PDMS substrate.

3.5 Improvements in wrinkle uniformity

3.5.1 Rotational Motion

A substantial improvement was yielded in wrinkle uniformity within a sample from the rotational motion setup compared to the

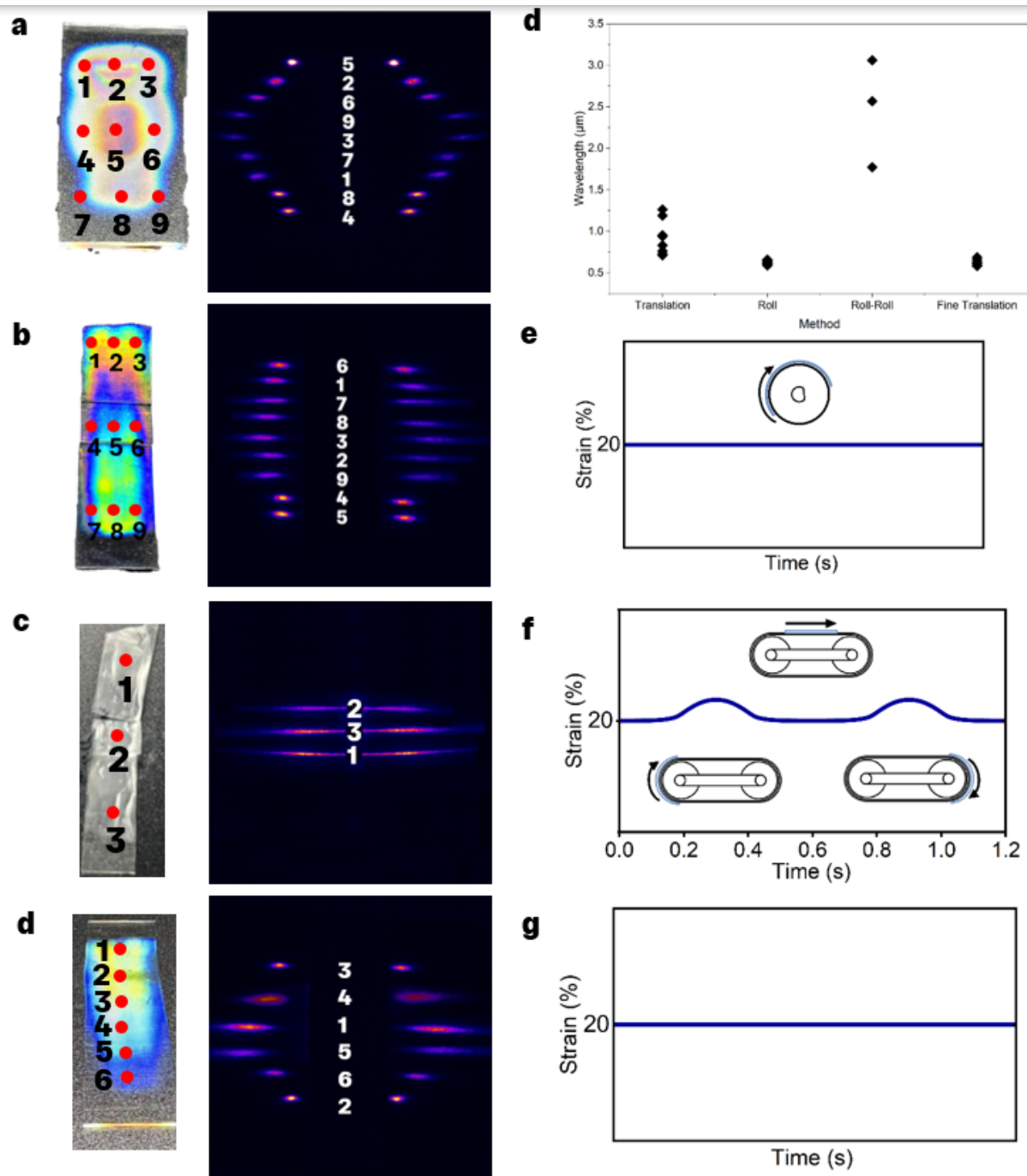


Figure 3 **a**, Sample from simple translational set-up showing variation in wavelengths, and points sampled across sample. Alongside is the SALS image showing variation in wavelengths across the 9 samples taken across the sample. **b**, Sample from rotational set-up showing variation in wavelengths, and points sampled across sample. Alongside is the SALS image showing variation in wavelengths across the 9 samples taken across the sample. **c**, Sample from roll-to-roll continuous set-up showing variation in wavelengths, and points sampled across sample. Alongside is the SALS image showing variation in wavelengths across the 9 samples taken across the sample. **d**, Sample from fine translational set-up showing variation in wavelengths, and points sampled across sample. Alongside is the SALS image showing variation in wavelengths across the 9 samples taken across the sample. **d**, Variation in wavelengths across samples compared across methods **e**, Graph showing variation of strain with time for rotational setup **f**, Graph showing variation of strain with time for roll-to-roll continuous processing setup **g**, Graph showing variation in strain for fine translation setup

simple translational motion setup. This is demonstrated by the standard deviation values of the 9 sampling points sampled which

was $0.03\mu\text{m}$ (wavelengths ranging from $0.59\mu\text{m}$ to $0.66\mu\text{m}$) as opposed to the $0.2\mu\text{m}$ value for the simple translational approach,

representing an 85% increase in uniformity by this metric. In Figure 3b, the 9 points distributed in a 3 x 3 grid across the sample surface provide a representative map of wrinkle wavelength uniformity, and are analysed using small-angle light scattering (SALS), as described in Section 2.4.2, which revealed a first-order spacing of points across the sample confirming spatial homogeneity in the wrinkle wavelength. By addressing the non-uniform butterfly-shaped exposure pattern inherent to the stationary atmospheric plasma module described in Section 3.1, the continuous rotation of pre-strained substrate during plasma exposure effectively averages the spatial distribution of plasma dose across the sample surface. Therefore, validating that continuous, symmetric exposure of the sample to the plasma jet can overcome intrinsic spatial non-uniformity of the atmospheric plasma intensity distribution. It is observed that the localized intensity gradients that would otherwise produce wavelength variations are smoothed out by continuously reorienting the substrate relative to the plasma source.

Optical microscopy, as discussed in Section 2.4.3, revealed that the sample treated with rotational motion exhibited a low density of surface cracks, wrapping the PDMS around the cylindrical surface remained below the critical threshold for mechanical failure¹⁴. For applications requiring continuous surface patterns with limited defects, the absence of significant cracking is particularly advantageous, as crack formation can compromise functional performance in applications such as controlled wetting by disrupting periodic wrinkle structure.

The rotational approach provides a scalable domain for applications dealing with cylindrical or tubular geometries, where the substrate naturally conforms to a rotating mandrel-like configuration. This approach provides a low-cost modification to enhance process uniformity, requiring minimal additional equipment beyond a cylinder (3-D printed in this case, or could be a drum for scale-ups), and a simple motor.

3.5.2 Roll-to-roll Motion: Continuous Processing

The roll-to-roll continuous processing approach was pursued to establish the feasibility of treating extended planar sample lengths in a conveyor-belt-like, scalable approach. Despite this architecture demonstrating the conceptual practicality of continuous, automated plasma treatment for industrial applications, critical mechanical challenges were revealed regarding strain maintenance that fundamentally limit the uniformity of wrinkle characteristics achievable in this configuration upon its implementation.

It was difficult to visualise the spatial distribution of wrinkle wavelengths across the roll-to-roll treated sample due to the presence of widespread cracks in the sample as shown by Figure 3.c, that disrupt the periodic wrinkle pattern, severely compromising the optical diffraction "fingerprint" visible in the SALS projection. The attainment of critical strain threshold for mechanical failure in the bilayer system causes these cracks. This emerges from the mechanical geometry of the roll-to-roll setup as shown in Figure 3.f. where: as the pre-strained PDMS sample transitions from the flat conveyor belt section to the curved roller section, and back again, the maintained strain undergoes pronounced cyclic variation.

A fundamental mechanical limitation of applying planar substrates to automated plasma treatment using roll-to-roll continuous processing is the geometric requirement that the pre-strained, flat substrate will eventually move around cylindrical rollers. This generates a non-uniform strain field that has proven incompatible with the strain uniformity demands of surface wrinkling applications.

3.5.3 Fine Translation Motion

The most significant enhancement in wrinkle uniformity within a sample was achieved by the fine translation motion setup, surpassing the rotational approach. This configuration successfully addressed the fundamental limitation of atmospheric plasma i.e. the inherent non-uniform butterfly-shaped exposure pattern characterised by stationary treatment, due to its exceptionally slow translational speed of 0.05 mm/s.

By sampling 6 points within a sample using small-angle light scattering (SALS) as shown in Figure 3., the distribution across the sample surface revealed a standard deviation in wrinkle wavelength of $0.04\mu\text{m}$, representing an 80% improvement over the simple translational setup ($0.20\mu\text{m}$). Majority of the wavelength measurements were clustered around $0.60\mu\text{m}$ and $0.65\mu\text{m}$, otherwise exhibiting a tight distribution ranging from $0.59\mu\text{m}$ to $0.67\mu\text{m}$. From the narrow spread, it can be confirmed that fine and continuous translational movement at 0.05 mm/s enabled each region of the sample received a highly consistent plasma dose, effectively averaging out the spatial intensity gradients inherent to the atmospheric plasma module. The speed was chosen to be 0.05 mm/s to strike a balance between dose uniformity and processing throughput. Slower speeds would increase processing time proportionally without necessarily improving uniformity, while also presenting the risk localised heating or oxidation that could increase crack formation. Conversely, the faster speeds would begin to reintroduce spatial heterogeneities as observed in the simple translational setup.

The glass-backed fine translation setup preserved uniform mechanical strain throughout the sample unlike the roll-to-roll configuration, which presented cyclic strain variations as samples transitioned around the cylindrical rollers. This stability in strain, coupled with the continuous and slow translational motion, ensured spatially uniform wavelengths and therefore glassy film thicknesses (as related by Equation 5), translating to uniform wrinkle characteristics upon strain release.

Cracks were revealed on the sample upon examination by optical microscopy. However, cracks were present to a lesser extent compared to the roll-to-roll configuration. This was thought to be attributed to the "flexing" motion undergone by the thin PDMS (thickness = $0.5\mu\text{m}$) when being released from the glass backing, that causes the glassy layer with the higher modulus³² of 40 GPa to fracture.

4 Conclusion

Atmospheric plasma-induced wrinkling of polydimethylsiloxane represents a practical and scalable alternative to the established ultraviolet-ozone and vacuum plasma oxidation techniques for nanostructured surface patterning. Our research demonstrated

that atmospheric plasma treatment generates wrinkle wavelengths and amplitudes comparable to those produced by conventional methods, while offering benefits for industrial implementation like in-line automation. The effects of plasma dosage and elastomer strain were studied to validate the frontal vitrification model previously developed for vacuum systems and establish fundamental mechanisms governing wrinkle formation.

The key finding of our research is that continuous motion configurations enhance spatial uniformity of wrinkles on a given PDMS sample, while also providing the scale-up benefits of automation. An 80% improvement was observed for the fine translational motion compared to the simple translational motion, reducing standard deviation of wavelengths from $0.2\ \mu\text{m}$ for the simple translational motion to $0.04\ \mu\text{m}$ for the fine translational motion. For the fine translational motion, tightly clustered wavelengths ranging between $0.58\ \mu\text{m}$ and $0.69\ \mu\text{m}$ were measured across the sample. This can be attributed to the $0.05\ \text{mm/s}$ translational motion effectively averaging the inherent non-uniform spatial intensity distribution of the atmospheric plasma jet. The rotational motion strategy showed similar results with an 85% improvement in wavelength uniformity within the sample from the simple translational motion, also representing a solution for cylindrical or tubular geometries while maintaining a low crack density. Conversely, roll-to-roll continuous processing, which was appealing as a concept for industrial translation, presented mechanical challenges in terms of cyclic strain variations as the sample cycles around rollers, inducing widespread cracking that disrupt periodic wrinkle structure and compromises functional performance.

It is established that pre-strain and atmospheric plasma dosage provide independent and tunable controls over wrinkle characteristics. The underlying diffusion-limited frontal oxidation process is shown by the logarithmic dependence of both amplitude and wavelength with exposure time, reflecting marked mechanistic similarities with vacuum systems despite the different exposure geometry. By varying the module-to-sample distance, we were able to consider an additional processing parameter otherwise unavailable in chamber-based systems. The established bilayer mechanics relationship was utilised to calculate film thicknesses from the wavelengths of these samples, producing values ranging from $2.26\ \text{nm}$ to $34.87\ \text{nm}$ with corresponding wavelengths spanning from $0\ \mu\text{m}$ (for an unwrinkled surface corresponding to 60 s exposure at a module-to-PDMS surface distance of 8mm) $3.24\ \mu\text{m}$ (for 3600 s of exposure at a module-to-PDMS surface distance of 2mm).

We conclude that these results position atmospheric plasma as a viable alternative to existing surface patterning methods in applications requiring large-area treatment. The high capital expenditure, energy consumption, and operational complexity associated with vacuum pump-down cycles and controlled environment chambers is eliminated by the use of atmospheric plasma devices which have shown to produce comparable results under atmospheric conditions. Being able to treat large or complex geometries without specialised infrastructure makes atmospheric plasma an attractive option for adoption. However, to reach full industrial potential, further optimisation is needed. The maxi-

mum translational speed for the fine translational setup that balances uniformity and industry standard throughput must be identified and reported. Parallel processing of multiple samples and wider atmospheric plasma modules must be tested and developed to expand treatment area in a single pass.

Finally, the demonstrated uniformity achievable through motion strategies, and the mechanistic understanding established here must be coupled to provide a foundation for successful industrial translation of atmospheric plasma as a surface patterning technology.

5 Experimental Section

PDMS elastomer films. Polydimethylsiloxane (PDMS) elastomers were prepared by mixing Sylgard 184 (Dow Corning) at 10:1 base to crosslinker ratio, and degassed under vacuum (10^{-2} mbar) for 20 min; Strips ($1.5\ \text{cm} \times 5\ \text{cm}$) were cut using a scalpel and any artefacts were removed using Scotch tape before use. **PDMS elastomer films.** Polydimethylsiloxane (PDMS) elastomers were prepared by mixing Sylgard 184 (Dow Corning) at 10:1 base to crosslinker ratio, and degassed under vacuum (10^{-2} mbar) for 20 min; Strips ($1.5\ \text{cm} \times 5\ \text{cm}$) were cut using a scalpel and any artefacts were removed using Scotch tape before use.

Atmospheric plasma oxidation. Piezobrush PZ23, max power 18W, with options for far-field (for non-conductive materials) and near-field module (for conductive materials)

Translational stages.

Rotational stages. A Zheng 12V DC gear-box motor (max. 200 rpm) was used to induce rotational motion for both the roll and R2R setup. The rpm could be altered by varying the voltage of the DC power supply. The cylindrical drums were custom-3D printed using Polylactic Acid (PLA) and the R2R apparatus consisted of a CO_2 laser cut base, ball bearings, and a silicone rubber wrist band.

Acknowledgements

We thank Zain Ahmed for his continuous experimental support and innovative contributions. We also thank Advanced Hackspace Imperial College London, for access to prototyping facilities (CO_2 Laser and 3D printers). Special mentions go to Hui Mao, Luis Torquato and Hisay Lama for their unwavering support

XX

References

- 1 D. S. Kim, H. W. Lee, J. H. Lee, H. G. Kwon, S. W. Lee, S. J. Han and O. C. Jeong, *Colloids and Surfaces B: Biointerfaces*, 2018, **170**, 266–272.
- 2 L. Pellegrino, L. S. Kriem, E. S. Robles and J. T. Cabral, *ACS Applied Materials & Interfaces*, 2022, **14**, 31463–31473.
- 3 A. Pirouz, I. Papakonstantinou and M. Michalska, *Frontiers in Chemistry*, 2024, **12**, 1354755.
- 4 Kommeren and As.
- 5 K. Khare, J. Zhou and S. Yang, *Langmuir*, 2009, **25**, 12794–12799.

- 6 L. Ma, L. He and Y. Ni, *Journal of Applied Physics*, 2020, **127**, 111101.
- 7 Z. Zheng, X. Gu, S. Yang, Y. Wang, Y. Zhang, Q. Han and P. Cao, *Biomimetics*, 2025, **10**, 470.
- 8 X.-M. Zhao, Y. Xia and G. M. Whitesides.
- 9 Y. Xia and G. M. Whitesides, *Angewandte Chemie International Edition*, 1998, **37**, 550–575.
- 10 Stensborg's Flexible Alternative to PhotoLithography for both Transparent and Non-transparent Materials: From Lab-Scale Soft Lithography to Industrial Roll-to-Roll Nanoimprint - Stensborg A/S, <https://stensborg.com/stensborgs-flexible-alternative-to-photolithography-for-both-transparent-and-non-transparent-materials-from-lab-scale-soft-lithography-to-industrial-roll-to-roll-nanoimprint/>.
- 11 Z. Zhang, L. Zhao and J. Shi, *Frontiers in Mechanical Engineering*, 2022, **8**, 910415.
- 12 S. Budday, S. Andres, Bastianwalter, P. Steinmann and E. Kuhl, *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 2017, **375**, 20160163.
- 13 S. Nikraves, D. Ryu and Y. L. Shen, *Scientific Reports* 2020 **10:1**, 2020, **10**, 5728–.
- 14 I. Miranda, A. Souza, P. Sousa, J. Ribeiro, E. M. Castanheira, R. Lima and G. Minas, *Journal of Functional Biomaterials*, 2021, **13**, 2.
- 15 L. B. Neves, I. S. Afonso, G. Nobrega, L. G. Barbosa, R. A. Lima and J. E. Ribeiro, *Micromachines*, 2024, **15**, 670.
- 16 I. Teixeira, I. Castro, V. Carvalho, C. Rodrigues, A. Souza, R. Lima, S. Teixeira, J. Ribeiro, I. Teixeira, I. Castro, V. Carvalho, C. Rodrigues, A. Souza, R. Lima, S. Teixeira and J. Ribeiro, *AIMS Materials Science* 2021 **6:952**, 2021, **8**, 952–973.
- 17 A. Müller, M. C. Wapler and U. Wallrabe, *Soft Matter*, 2019, **15**, 779–784.
- 18 A. M. Smith, D. G. Inocencio, B. M. Pardi, A. Gopinath and R. C. A. Eguiluz, *ACS Applied Polymer Materials*, 2024, **6**, 2405–2416.
- 19 Z. Ahmad, B. P. M. R, H. Barr and J. T. Cabral, *Nano Letters*, 2025, **25**, 740–746.
- 20 Plasma Treatment for PDMS Bonding — Tergeo plasma cleaner, <https://piescientific.com/resource-pdms-bonding/>.
- 21 A. Zeniou, D. Kefallinou, P. Dimitrakellis, E. Xenogiannopoulou, M. Grigoriou, A. Dimoulas, D. T. Boumpas, A. Tserepi and E. Gogolides, *Chempluschem*, 2024, **89**, e202400290.
- 22 M. Nania, O. K. Matar, J. Ao and T. Cabral, 2015.
- 23 A. D. Campo, A. Nogales, T. A. Ezquerria and J. Rodriguez-Hernández.
- 24 How to Overcome Challenges in Plasma Processing, <https://www.surfxttechnologies.com/blog/how-to-overcome-challenges-in-plasma-processing/>.
- 25 S. M. Hong, S. H. Kim, J. H. Kim and H. I. Hwang, *Journal of Physics: Conference Series*, 2006, **34**, 656–661.
- 26 Roll to Roll - Plasmalex, <https://www.plasmalex.com/roll-to-roll/>.
- 27 Atmospheric Plasma Treatment Applications Notes | INTRFACE TECHNOLOGIES, <https://www.intrfacetech.com/atmospheric-plasma-treatment-applications-notes/>.
- 28 H. T. Sasmazel, M. Alazzawi and N. K. A. Alsahib, *Molecules*, 2021, **26**, 1665.
- 29 D. Korzec, F. Hoppenthaler, D. Burger, T. Andres and S. Nettesheim, *Plasma Processes and Polymers*, 2020, **17**, 2000053.
- 30 C. Yu and H. Jiang, 2010.
- 31 X. Cheng, L. Miao, Z. Su, H. Chen, Y. Song, X. Chen and H. Zhang, *Microsystems & Nanoengineering* 2017 **3:1**, 2017, **3**, 16074–.
- 32 F. A. Bayley, J. L. Liao, P. N. Stavrinou, A. Chiche and J. T. Cabral, *Soft Matter*, 2014, **10**, 1155–1166.
- 33 J. Song, H. Jiang, Z. J. Liu, D. Y. Khang, Y. Huang, J. A. Rogers, C. Lu and C. G. Koh, 2008.
- 34 Z. Ahmad, B. Parias M. R, H. Barr and J. T. Cabral, *Nano Letters*, 2024, **25**, 740–746.

Noble-Metal-Free CdS/MoP Photocatalyst for Enhanced Selective Conversion of Methanol to Ethylene Glycol under Visible Light

Tingxu Zhao, Yuye Zhang

Department of Chemical Engineering, Imperial College London, United Kingdom

Abstract

Ethylene glycol (EG) is a crucial chemical in industrial production such as plastics. The global production of EG has exceeded 40 Mt/yr while it's an energy intensive process with massive carbon dioxide (CO₂) emission. Photocatalytic methanol to EG (MTEG) reaction has been attractive due to its sustainability and considered as a “dream catalytic reaction”. However, it is still limited by the low selectivity and conversion. To overcome this, this paper proposes a novel MoP cocatalyst loaded onto CdS photocatalyst to enhance selectivity and conversion rate of EG in MTEG reaction under visible light via C-C coupling. The synthesis of cocatalyst involved calcination of precursors at varied temperatures. The XRD and SEM results showed MoP was not synthesised, but a mixture of MoO₂ and MoO₃ was formed and still improved the reaction activity. The optimum catalytic performance occurred on the cocatalyst synthesised at 700°C with 3 wt% loaded onto CdS, that increased selectivity and conversion rate by 40% and 68% respectively compared with base CdS catalyst.

Keywords

Photocatalysis, C-C Coupling, Molybdenum Phosphide, Cadmium Sulphide, Methanol, Ethylene Glycol

1. Introduction

Ethylene glycol (HOCH₂CH₂OH, EG) is a key industrial chemical widely used in the production of plastics, antifreeze, polyester fibres, as well as a fundamental building block for various downstream chemicals. [1] The global production of ethylene glycol exceeded 40 Mt in 2020 and is expected to increase by 5–10% annually. [2]

Traditionally, ethylene glycol is produced through a highly energy-intensive two-step process in integrated chemical plants operated under high temperatures and pressures with substantial CO₂ emissions and other waste streams. In this process, ethylene is first oxidized by oxygen (O₂) on a silver-based catalyst to form ethylene oxide, followed by a hydrolysis reaction of ethylene oxide to ethylene glycol catalysed by strong acid. The overall product selectivity is below 80% for this process due to the inevitable overoxidation of ethylene in the first step. [2] These limitations have motivated increasing interest in sustainable routes for EG synthesis based on abundant, low-toxicity C₁ feedstocks.

many reaction processes, especially for the challenging reactions containing selective C–H activation and C–C coupling in chemical synthesis. The direct coupling of methanol to ethylene glycol (MTEG) has been attractive since it could be achieved via solar energy. [3] Methanol is a clean platform chemical, which can not only conventionally be produced from natural gas and coal, but also has been directly synthesised from biomass and CO₂. [4] Therefore the visible light-driven MTEG route provides an alternative process for sustainable synthesis of EG.

Among various semiconductors, CdS has demonstrated an outstanding photocatalytic performance in MTEG reaction compared to other metal sulphides and oxides (Table 1). [5] However, the methanol conversion efficiency of unary CdS semiconductors is still limited due to the rapid recombination of electron-hole pairs and photo corrosion. To overcome this limitation, co-catalyst is introduced to improve charge separation and consequently catalytic efficiency.

This paper presents a novel CdS/MoP photocatalyst for the selective conversion of methanol to ethylene

Photochemistry has been applied in controlling

glycol. By integrating CdS nanorods with a noble-metal-free MoP cocatalyst, the study seeks to

improve ethylene glycol selectivity and yield.

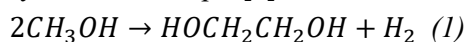
Table 1 Catalytic performances of typical semiconductors in MTEG [5]

Catalyst	Reaction condition	EG Formation rate (mmol g _{cat} ⁻¹ h ⁻¹)	Selectivity (%)
ZnS	UV-Vis light	1.3	54
ZnS	Visible light	0	0
Cu ₂ O		0	0
CdS particle		0.28	58
CdS rod		0.46	71

2. Background

2.1. Reaction mechanism

The methanol to ethylene glycol (MTEG) reaction (Eq.1) is a dehydrogenative C-C coupling reaction when the C-H bond of methanol is activated. It basically includes 3 steps, [5] as illustrated in Fig.1



First a photon is absorbed and promotes an electron from the valence band to the conduction band of photocatalyst, generating an electron-hole pair. The photogenerated hole can then activate the C-H bond of a methanol molecule absorbed on the catalyst surface. Two pathways have been proposed for this C-H activation. In the first, a stepwise proton-transfer followed by electron-transfer occurs. Heterolytic cleavage of the C-H bond forms CH_2OH^- and a proton, then electron transfer from CH_2OH^- to the photogenerated hole, yielding a hydroxymethyl ($\text{CH}_2\text{OH}\cdot$) radical. Alternatively, C-H activation may proceed through homolytic cleavage in a concerted proton-electron transfer (CPET) step, directly producing the $\text{CH}_2\text{OH}\cdot$ radical. Finally, two $\text{CH}_2\text{OH}\cdot$ radicals couple to form ethylene glycol.

proton transfer to form a $\text{CH}_3\text{O}\cdot$ radical and consequently be oxidised to formaldehyde (Eq.2).



Generally, the O-H bond in methanol is naturally weaker than C-H bond and more easily activated, making formaldehyde a preferable product which inevitably limits EG selectivity. Catalysts achieving high selectivity towards EG therefore requires promoting C-H activation over O-H activation. Based on proposed reaction pathway, [5] the photocatalytic performance in MTEG reaction is affected by two factors. One is the efficiency of electron-hole pair separation, which reflects the degree of charge recombination and the other is photocatalyst band gap, which controls the ease of photoexcitation and the fraction of visible light photons that are capable to drive the reaction. Maximising EG selectivity and yield is essential for efficient utilisation of methanol and simplified following separation process. Consequently, catalytic systems applied to this reaction must optimise these two parameters.

2.2. CdS catalyst

To achieve successful catalysis in MTEG reaction, the conduction band of the catalyst must be more negative than the reduction potential of methanol to EG (-0.38 V at pH 7), while the valence band must be more positive than the reduction potential of H^+/H_2 . As the potential required for C-H activation is close to that of hydrogen evolution, a range of HER catalyst have been explored (Table 1) with their band gaps presented in Fig.2.

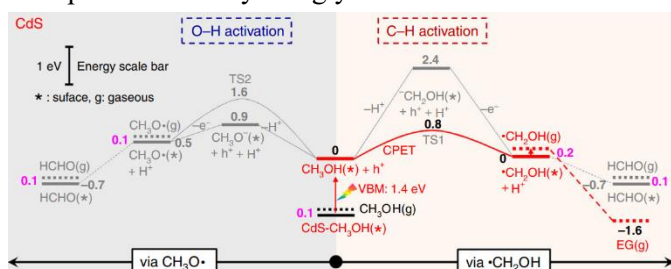


Fig.1 MTEG reaction pathway with CdS photocatalyst proposed by Xie et al. [5]

However, as shown in Fig.1, C-H activation competes with O-H activation, which proceeds via

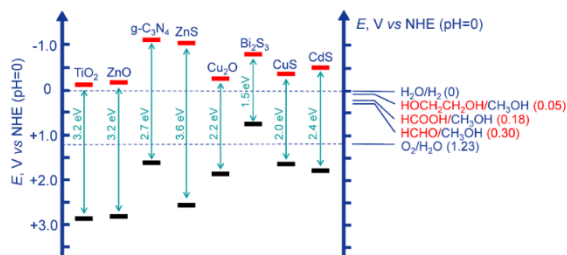


Fig.2 Band gaps of typical HER semiconductors [5]

From the catalytic performance illustrated by Table 1 and Fig.2, CdS is considered as a particularly suitable photocatalyst due to its relatively narrow band gap (2.4 eV) and excellent charge capacity, both of which enhance the electron-hole separation. [6] In addition, the morphology of CdS plays a significant role in determining catalytic performance. CdS nanorods exhibits exceptional high activity and EG selectivity than other morphologies. [5]

2.3. MoP cocatalyst

To improve photocatalyst performance, co-catalysts are introduced to modify the semiconductors. This improvement is achieved by increasing of electron-hole separation efficiency, suppression of charge recombination or modification of the band gap of the system. [7]

Molybdenum phosphide (MoP), an earth-abundant and non-noble transition metal phosphide, has been found to display good electronic conductivity, superior chemical stability and platinum (Pt)-like electronic structure and extensively studied as a promising replacement for platinum in the hydrogen evolution reactions. [8] In photocatalytic hydrogen evolution, MoP has been shown to significantly improve the performance of CdS and g-C₃N₄ by suppressing electron-hole recombination and accelerating interfacial reduction kinetics [9,10]. Particularly, MoP forms intimate heterojunctions in CdS that facilitate rapid electron extraction from the semiconductor, remarkably increasing quantum efficiency under visible light. Similar enhancements have also been observed in MoP-modified g-C₃N₄, red phosphorus, and SiC systems, where MoP promotes spatial charge

separation and provides low-barrier pathways for multi-electron surface reactions [10-12]. Thus, MoP represents a cost-effective and attractive co-catalyst in such photocatalytic reaction.

However, whether this outstanding catalytic performance of MoP co-catalyst could transfer to other similar reactions such as C-C coupling reaction remains unclear, although methanol-to-ethylene-glycol (MTEG) reaction pathway also requires efficient extraction of photogenerated electrons to drive methanol C-H activation while suppressing competing O-H activation. Noting this research gap, this paper proposes a novel CdS/MoP photocatalyst to enhance both activity and selectivity in the MTEG reaction.

3. Methodology

3.1 Synthesis of CdS nanorods

CdS nanorods were synthesised through a solvothermal method. [13] In this procedure 4.62 g of cadmium chloride hemi-pentahydrate and 4.62 g of thiourea were dispersed in 60 ml of ethylenediamine. The mixture was then transferred to a Teflon-lined autoclave and was maintained at 160 °C for 24 h. After cooling down to room temperature, the yellow solid product was collected by centrifugation, washed with distilled water and ethanol three times, then dried at 60 °C overnight.

3.2. Synthesis of MoP co-catalyst

To prepare MoP cocatalyst, 0.618 g (NH₄)₆Mo₇O₂₄·4H₂O (0.5 mmol) and 0.420 g NaH₂PO₂·H₂O (3.5 mmol) were ground in an agate mortar to well blend. Then the acquired well-mixed mixtures were placed in a quartz boat and calcined in a tube furnace at 300°C under N₂/H₂ atmosphere for 5 h (ramping rate, 5 °C min⁻¹). When cooled to room temperature, the product was collected and washed three times with deionised water. Then the samples from the previous step were dried at 60°C to obtain the final product and labelled MoP (300°C). [14]

3.3. Loading of cocatalyst onto catalyst

The MoP/CdS photocatalyst was synthesised using an ultrasonication method. [15] In a typical procedure, 10 mg of MoP was first ultrasonically dispersed in 10 ml N, N-dimethylformamide (DMF) in a centrifuge tube for 3 h at room temperature. To achieve desired loading rates, variable amounts of this suspension were added to a separate centrifuge tube with 100 mg of CdS and topped up to 10 ml with fresh DMF, following which this new mixture was dispersed ultrasonically for 2 h to achieve close contact between MoP and CdS. The loading amount of MoP onto CdS was varied from 1 w/t% to 5 w/t%, calculated as a nominal percentage of the mass of the CdS added. The obtained products were washed with deionised water and ethanol twice each, then dried in a vacuum oven at 60 °C for 24h.

3.4. Photoreactor experiments

Photocatalytic reactions were carried out in an EvoluChem Photo Redox Box high throughput reactor, shown in Fig. 2 below, which can hold up to 4 reaction vials at a time. The reactor utilises a 450 nm, 30 W LED lamp as light source. To prepare each reaction vial, 10 mg of solid catalyst was mixed with 4.0 cm³ methanol and 1.0 cm³ deionised water, with a magnetic stirrer added to each vial to ensure sufficient mass transfer. Each vial was then evacuated by purging the vial with N₂ gas using a Schlenk line, then loaded into the reactor (Fig. 3). Photocatalytic reactions were performed for 20h under magnetic stirring at 700 rpm. Following the reaction, the composition of liquid products was analysed using a high-performance liquid chromatograph (HPLC, Shimazu LC-20A) with refractive index and ultraviolet detectors. A Supelcogel C-610H column (30 cm by 7.8 mm) and a 0.005M sulfuric acid as the mobile phase were used.



Fig.3 Photocatalytic Reactor Setup

3.5. Characterisation techniques

To characterise the material, X-ray diffraction (XRD) and scanning electron microscopy (SEM) were conducted to investigate the morphology of the catalyst. XRD patterns were measured on a Malvern Panalytical XRD X'Pert PRO system using, while field-emission gun SEM micrographs were obtained using a LEO Gemini 1525 FEG-SEM and an electron high tension of 5 kV.

4. Results

4.1 SEM and XRD results

4.1.1 Cocatalyst synthesised at 300°C

XRD and SEM analysis were conducted to confirm whether MoP is successfully synthesised at 300°C. X-ray diffraction (XRD) patterns were collected using Cu K α radiation ($\lambda = 1.5406 \text{ \AA}$). Phase identification was performed by comparing the pattern of sample with reference images of possible products from the Inorganic Crystal Structure Database (ICSD). As shown by Fig5a, the XRD pattern of 300°C sample exhibits strong diffractions at $2\theta = 23.7^\circ, 25.8^\circ, 27.4^\circ, 33.7^\circ$ and 39.2° that match the reference lines of MoO₃ instead of MoP

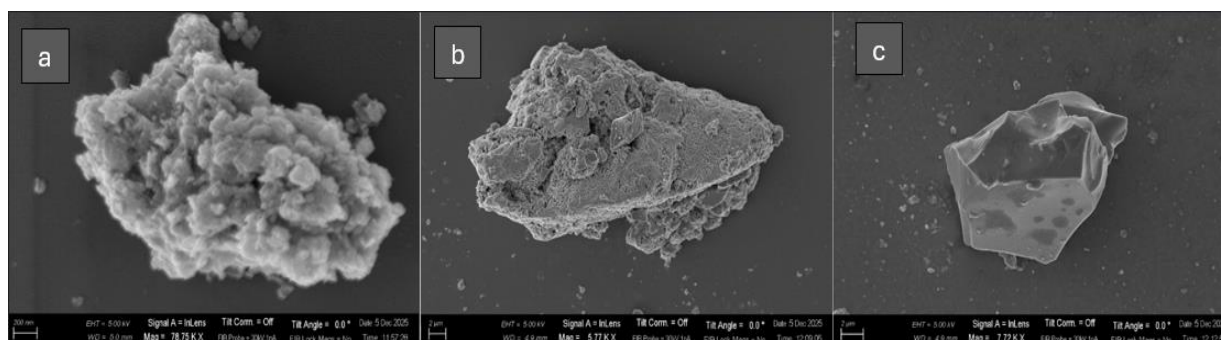


Fig. 4a-c: SEM of a) 300°C sample, b) 600°C sample, and c) 700°C sample

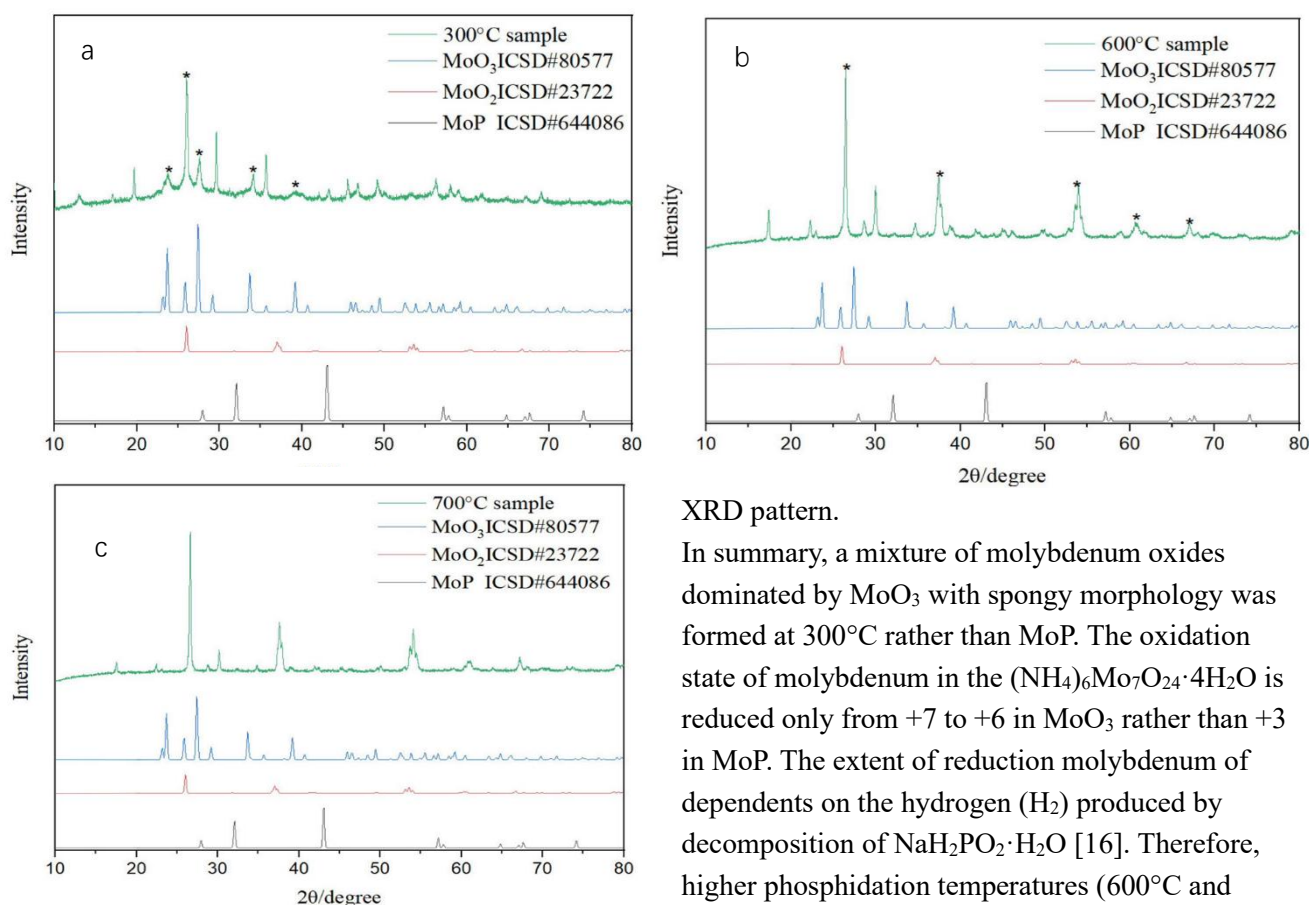


Fig. 5a-c: XRD of a) 300°C sample, b) 600°C sample and c) 700°C sample

WC type. Only a few weak peaks at 37° and 53.6° match the reference lines of MoO₂, indicating 300°C sample is mainly composed by MoO₃ with minor amount of MoO₂. Additionally, XRD pattern shows a high noise level. SEM images reveal morphology at the nano/microscale and provide qualitative clues about crystallinity. The SEM image of 300°C sample (Fig. 4a) shows rough and spongy-like nanoparticle aggregates with no discernible faceting that confirms the noisy peaks in

XRD pattern.

In summary, a mixture of molybdenum oxides dominated by MoO₃ with spongy morphology was formed at 300°C rather than MoP. The oxidation state of molybdenum in the (NH₄)₆Mo₇O₂₄·4H₂O is reduced only from +7 to +6 in MoO₃ rather than +3 in MoP. The extent of reduction molybdenum depends on the hydrogen (H₂) produced by decomposition of NaH₂PO₂·H₂O [16]. Therefore, higher phosphidation temperatures (600°C and 700°C) were then applied for MoP synthesis to promote H₂ production and consequently the reduction of molybdenum.

4.1.2 Cocatalyst synthesised at 600°C and 700°C

Strong diffractions at 2θ = 26°, 37°, 53.6°, 60.5° and 66.7° in XRD image of 600°C sample (Fig. 5b) correspond to MoO₂ pattern. Although MoP did not form, this increased degree of reduction confirms the necessity of high temperatures for MoP synthesis. Compared with 300°C sample, there are less noisy features appeared at the XRD pattern of

600°C sample, showing increased crystallinity. SEM image of 600°C sample (Fig.4b) confirms enhanced crystallization and particle growth. Similar to 600°C sample, a smoother MoO₂ pattern appeared on the XRD image for 700°C sample (Fig.5c). Thin, faceted, crystal-like flakes with sharp edges are observed from the SEM image of 700°C sample (Fig.4c), which is consistent with lower noise level in its XRD pattern.

4.2 Results of MTEG reaction

4.2.1 CdS/MoO₃ (300°C)

From Fig. 6 with increasing of MoO₃(300°C) loading from 1 to 3 wt%, the conversion rate improves substantially and peaks at 2wt% with 2.7 mmol·g_{cat}⁻¹·h⁻¹ that is almost 7 times higher than that of base CdS. While further increasing the loading rate to 3wt%, the conversion rate drops sharply to 0.86 mmol·g_{cat}⁻¹·h⁻¹. By contrast, the selectivity was not improved as the highest selectivity (37%) reached at 3wt% is almost the same as that of CdS.

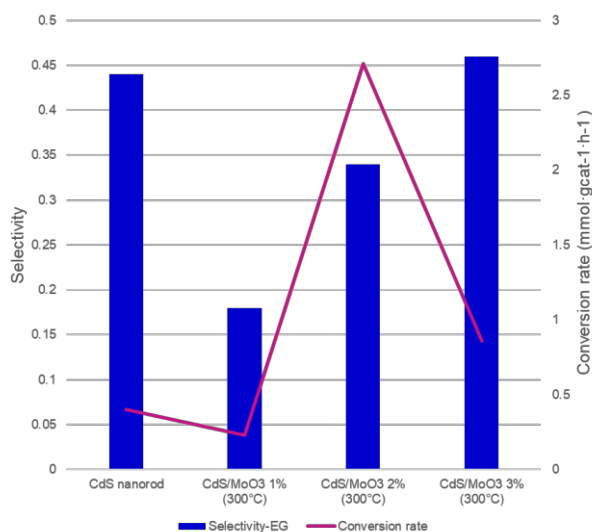


Fig.6 CdS/MoO₃ (300°C) MTEG reaction results

4.2.2 CdS/MoO₂ (600°C) and CdS/MoO₂ (700°C)

As presented by Fig. 7 CdS/MoO₂ (700°C) series has a better performance than that of CdS/MoO₂ (600°C) series for both EG selectivity and conversion rate. This result suggests that the maximum selectivity and conversion rate was achieved in CdS/MoO₂ 3% (700°C) at the same time, with 62% and 0.67 mmol·g_{cat}⁻¹·h⁻¹ respectively. Similarly, once the loading rate

exceeds this optimum point, a rapid decreasing occurred. While for the CdS/MoO₂ (600°C) series, both selectivity and conversion rate are lower than those of CdS nanorod.

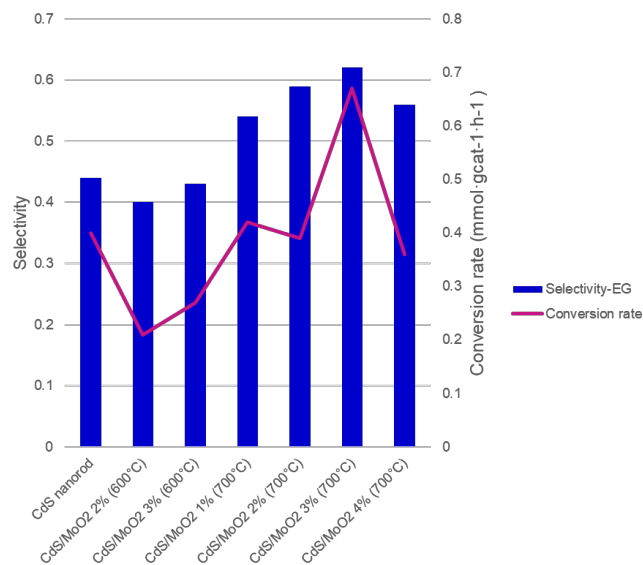


Fig. 7 CdS/MoO₂ (600°C) and CdS/MoO₂ (700°C)

MTEG results

5. Discussion and Conclusion

5.1 MoP synthesis

Formation of MoP is a continuous reduction process of (NH₄)₆Mo₇O₂₄·4H₂O. The unsuccessful synthesis of MoP was led by the incomplete decomposition of NaH₂PO₂·H₂O to generate enough H₂ driving the reduction reaction. Due to the safety concern, the highest calcination temperature is limited by 700°C while still remaining incapable to promote the complete reduction of Mo in (NH₄)₆Mo₇O₂₄·4H₂O precursor. To achieve the successful synthesis of MoP, calcination temperature is required to reach 800°C reported by previous literature. [14]

5.2 Catalytic performance of MoO_x

Although MoP was not formed as expected, the resulting MoO_x presented the functionality as a cocatalyst in MTEG reaction. By combining the experimental data analysis with the XRD and SEM characterisation techniques, the conclusions regarding MoO_x cocatalyst are as follows. The CdS/MoO₃ (300°C) enhanced the conversion rate while EG selectivity

was not improved. This is due to the heterojunction structure between CdS and MoO₃. The valence band maximum (VBM) of MoO₃ is around +2.7 eV which is higher than that of CdS (+1.9 eV), [17] beneficial for oxidation reactions. Thus, the conversion rate for both EG and formaldehyde in MTEG reaction increases while selectivity of EG remains unchanged. Due to high electric conductivity of MoO₂, electrons in catalytic reaction were absorbed to suppress the charge recombination. [18] Highest conversion rate was

therefore achieved in CdS/MoO₂ 3% (700°C), increasing 68% over base CdS catalyst. High selectivity should be confirmed via ¹³C-labelling with ¹³CH₃OH, checking whether the C-H activation is preferable over O-H activation. For MoO₂ (600°C), the unsatisfactory performance in MTEG reaction is mainly due to the lack of large surface area and active sites similar to those of MoO₂ (300°C) and MoO₂ (700°C).

Reference

- [1] Staples, C. A., Williams, J. B., Craig, G. R., & Roberts, K. M. (2001). Fate, effects and potential environmental risks of ethylene glycol: a review. *Chemosphere*, 43(3), 377–383. [https://doi.org/10.1016/s0045-6535\(00\)00148-x](https://doi.org/10.1016/s0045-6535(00)00148-x)
- [2] Fan, L., Zhao, Y., Chen, L., Chen, J., Chen, J., Yang, H., Xiao, Y., Zhang, T., Chen, J., & Wang, L. (2023). Selective production of ethylene glycol at high rate via cascade catalysis. *Nature Catalysis*, 6(7), 585–595. <https://doi.org/10.1038/s41929-023-00977-6>
- [3] Wang, L., Du, D., Zhang, B., Xie, S., Zhang, Q., Wang, H., & Wang, Y. (2021). Solar energy-driven C–H activation of methanol for direct C–C coupling to ethylene glycol with high stability by nitrogen doped tantalum oxide. *CHINESE JOURNAL OF CATALYSIS (CHINESE VERSION)*, 42(9), 1459–1467. [https://doi.org/10.1016/s1872-2067\(21\)63797-x](https://doi.org/10.1016/s1872-2067(21)63797-x)
- [4] Wang, M., Liu, M., Lu, J., & Wang, F. (2020). Photo splitting of bio-polyols and sugars to methanol and syngas. *Nature Communications*, 11(1), 1083. <https://doi.org/10.1038/s41467-020-14915-8>
- [5] Xie, S., Shen, Z., Deng, J., Guo, P., Zhang, Q., Zhang, H., Ma, C., Jiang, Z., Cheng, J., Deng, D., & Wang, Y. (2018). Visible light-driven C–H activation and C–C coupling of methanol into ethylene glycol. *Nature Communications*, 9(1), 1181. <https://doi.org/10.1038/s41467-018-03543-y>
- [6] Zheng, X., Liu, Y., Yang, Y., Song, Y., Deng, P., Li, J., Liu, W., Shen, Y., & Tian, X. (2023). Recent advances in cadmium Sulfide-Based photocatalysts for photocatalytic hydrogen evolution. *Renewables*, 1(1), 39–56. <https://doi.org/10.31635/renewables.022.202200001>
- [7] Fan, Y., Bao, J., Shi, L., Li, S., Lu, Y., Liu, H., Wang, H., Zhong, L., & Sun, Y. (2018). Photocatalytic Coupling of Methanol and Formaldehyde into Ethylene Glycol with High Atomic Efficiency. *Catalysis Letters*, 148(8), 2274–2282. <https://doi.org/10.1007/s10562-018-2465-6>
- [8] Xiao, W., Zhang, L., Bukhvalov, D., Chen, Z., Zou, Z., Shang, L., Yang, X., Yan, D., Han, F., & Zhang, T. (2020). Hierarchical ultrathin carbon encapsulating transition metal doped MoP electrocatalysts for efficient and pH-universal hydrogen evolution reaction. *Nano Energy*, 70, 104445. <https://doi.org/10.1016/j.nanoen.2020.104445>
- [9] Yin, S., Han, J., Zou, Y., Zhou, T., & Xu, R. (2016b). A highly efficient noble metal free photocatalytic hydrogen evolution system containing MoP and CdS quantum dots. *Nanoscale*, 8(30), 14438–14447. <https://doi.org/10.1039/c6nr00989a>
- [10] Tang, J., Yang, D., Zhou, W., Guo, R., Pan, W., & Huang, C. (2019). Noble-metal-free molybdenum phosphide co-catalyst loaded graphitic carbon nitride for efficient photocatalysis under simulated irradiation. *Journal of Catalysis*, 370, 79–87. <https://doi.org/10.1016/j.jcat.2018.12.009>

- [11] Ge, Z., Wang, M., Li, R., Jin, X., & Chen, Y. (2023). Nanosized MoP as a cocatalyst over red phosphorus for photocatalytic water splitting. *Materials Letters*, 340, 134185. <https://doi.org/10.1016/j.matlet.2023.134185>
- [12] Wang, W., Zhang, Y., Chang, H., Hu, S., Qiu, R., Zhang, M., Li, C., Li, Y., Wu, L., Yang, W., Jing, X., Xie, H., Liu, Q., Wang, X., & Crittenden, J. (2024). Revolutionary amorphous MoP quantum dots as efficient visible light harvester coupled with SiC for boosting photocatalytic hydrogen production. *Renewable Energy*, 235, 121389. <https://doi.org/10.1016/j.renene.2024.121389>
- [13] Jiang, D., Sun, Z., Jia, H., Lu, D., & Du, P. (2015). A cocatalyst-free CdS nanorod/ZnS nanoparticle composite for high-performance visible-light-driven hydrogen production from water. *Journal of Materials Chemistry A*, 4(2), 675–683. <https://doi.org/10.1039/c5ta07420g>
- [14] Ma, X., Li, W., Li, H., Dong, M., Li, X., Geng, L., Fan, H., Li, Y., Qiu, H., & Wang, T. (2022). Fabrication of novel and noble-metal-free MoP/In₂S₃ Schottky heterojunction photocatalyst with efficient charge separation for enhanced photocatalytic H₂ evolution under visible light. *Journal of Colloid and Interface Science*, 617, 284–292. <https://doi.org/10.1016/j.jcis.2022.03.021>
- [15] Zhang, H., Xie, S., Hu, J., Wu, X., Zhang, Q., Cheng, J., & Wang, Y. (2020). C–H activations of methanol and ethanol and C–C couplings into diols by zinc–indium–sulfide under visible light. *Chemical Communications*, 56(12), 1776–1779. <https://doi.org/10.1039/c9cc09205f>
- [16] Shirman, R. and Sasson, Y. (2023) 'Hydrogen generation from sodium hypophosphite catalyzed by metallic nanoparticles supported on graphitic carbon nitride,' *International Journal of Hydrogen Energy*, 48(71), pp. 27611–27618. <https://doi.org/10.1016/j.ijhydene.2023.03.434>.
- [17] Egbo, K.O. et al. (2021) 'Band alignment of wide bandgap NiO/MoO₃ and NiO/WO₃ p-n heterojunctions studied by high-resolution X-ray photoelectron spectroscopy,' *Journal of Alloys and Compounds*, 876, p. 160136. <https://doi.org/10.1016/j.jallcom.2021.160136>.
- [18] Shaikh, A.A. et al. (2024) 'A comprehensive review of the oxidation states of molybdenum oxides and their diverse applications,' *Sustainable Chemistry for the Environment*, 7, p. 100125. <https://doi.org/10.1016/j.scenv.2024.100125>.

Repeatability of ionic liquid treatment for cotton and dye recycling process

Nadia Tholley and Uma Keomany

Department of Chemical Engineering, Imperial College London, U.K.

Abstract The textile industry faces increasing environmental pressures due to intensive water usage, reliance on hazardous chemicals, and the low recycling rates of wastes. Colour removal and dye recovery remain key barriers to circular textile systems, particularly for cotton dyed with reactive dyes. This study investigates a process using a low-cost ionic liquid (IL), N,N-dimethylbutylammonium hydrogen sulfate [DMBA][H₂SO₄], to strip dyes from cotton while enabling IL and dye reuse, offering a promising solution for future separation and recycling of cotton. Across multiple treatment cycles, an average cotton yield of **91.8%** was achieved, demonstrating minimal fibre loss and promising recyclability. Wool dye uptake following IL-assisted dye recovery was 88%, confirming the potential for dye reuse in subsequent dyeing applications. While the process demonstrates feasibility, several areas require optimisation, including improving IL recovery efficiency, minimising dye degradation, and evaluating broader sustainability and economic impacts.

1. Introduction

The textile industry is responsible for mass amounts of waste and pollution annually. Between fabric waste ending up in landfill, large amounts of water usage in dye processes, pollution of toxic dyes to the environment and mass amounts of greenhouse gases produced in these processes, the effects of the fashion industry are extremely destructive and unsustainable. This has only worsened with the increase in fast fashion over the last few decades. In the UK alone, 350,000 tonnes of clothing is sent to landfill each year [1]. Currently, the fashion industry produces approximately 92 million tonnes of textile waste [2] and 79 billion cubic meters of water annually [1]. This is only expected to increase, with a predicted annual textile consumption of 102 million tonnes by 2030. The main cause of pollution in the fashion industry is the dyeing and finishing process which accounts for 36% of the industry's emissions [3]. Cotton is the most widely used natural fibre in the textile industry [8], having a global fibre production of 24.5 million tonnes in 2024 [9]. Consequently, the cotton industry is largely responsible for this environmental damage, with synthetic fertilisers and pesticides used in the cultivation process resulting in water contamination and greenhouse gas emissions. Within the wastewater produced from textile dyeing, harmful chemicals such as aromatic amines, heavy metals and synthetic compounds such as formaldehyde and other surfactants can be found [6] [4]. These pose significant risk to human and aquatic life due to the presence of carcinogens and toxic substances which are then released to the environment to poison water sources and soil. In addition to this, the heavy metals in the wastewater, such as lead, chromium and mercury from dye impurities, can alter soil properties and be taken up into crops where they cause further damage upon ingestion [5]. It is therefore vital to explore ways to reduce the use and production of dyes. Due to this, there has been a strong interest in recent years to reduce environmental burden through finding technologies to reduce the chemicals needed in fashion production and find ways to recycle the discarded fabrics and dyes.

One process that can be applied to utilise the tonnes of textiles wasted and reduce the need for the uses of new

dyes and fabrics, is extracting and recycling used dyes with a solvent. This process has been underutilised thus far due to drawbacks such as not being able to completely remove dye from cotton due to difficulty in finding capable solvents. This can give rise to unpredictable and inconsistent colouring of recycled fabrics which is unacceptable in the commercial fashion industry. In attempt to combat these issues, a band of solvent that are being investigated for this process are ionic liquids (ILs). By using ILs to remove the dye, the textiles can be redyed and recycled, for example via wet spinning for cotton fibres [7], thus increasing the lifetime of fabrics and reducing their production demand. The separated dyes can also be reused which reduces dye waste and the need for dye production. The process of reusing the dyes and fabric should theoretically be less energy and resource intensive than producing new textiles from raw materials, thus reducing carbon and pollution emissions from the fashion industry.

Polycotton is an abundant fabric blend, which is not recycled enough due to the difficulty in separating the cellulosic cotton from the synthetic polyester. One reason that cotton fibre hydrolysis is useful in fibre recycling because it allows the separation and degradation of polycotton, as the cellulosic cotton will be broken down while the polyester fibres will remain intact leading them to be easily separable. Cotton hydrolysis can be achieved under alkaline conditions or in acidic conditions, which occurs in DMBA[HSO₄]. The three step mechanism of acid hydrolysis involves first the protonation of the β -1,4 glycosidic bond. Following this the C-O bond is cleaved and then finally, the carbocation is attacked by a water molecule. This then produces β -glucose monomers in the form of powdered cotton to be regenerated into new fibres.

The aim of this project is to explore and optimise the conditions of reactive dye removal from cotton fabric using DMBA [H₂SO₄] as a stripping solvent and look into the feasibility of reusing the stripped dyes to colour wool fabric. Due to damage to polymers causing a change in the dyeability of recycled fibres being a previous drawback of the process, it is an objective to investigate the condition of the stripped cotton and the

extracted reactive dye in order to assess the effect of ILs on cotton and dye quality and degradation. Ensuring that the maximum recovery of ionic liquid is achieved is also important to reduce waste of materials and make the process as sustainable and economical as possible.

2. Background

2.1 Ionic Liquids

Cellulose is challenging to dissolve with conventional solvents due to its strong hydrogen bonding and high crystallinity [9], which limits fibre recyclability and contributes to environmentally damaging practices such as the viscose process. Ionic liquids (ILs) are defined as ionic salts that have a melting point lower than 100°C. It offers a promising alternative due to their low vapour pressure, non-flammability. Other properties such as their high thermal stability and strong solvation capabilities make them particularly attractive for dissolving cellulose and for processes involving elevated temperatures. However, the commercial use of classical aprotic ILs has been hindered by their high production cost. In contrast, protic ionic liquids (PILs) provide a low-cost pathway to IL-based cellulose processing, as they are produced via a simple proton-transfer reaction between a Brønsted acid and base, enabling large-scale synthesis from inexpensive precursors.[11]

2.2 Reactive dyes

Cotton is most commonly coloured with either direct dyes or reactive dyes. Direct dyes work by diffusing over and adsorbing to fibres whereas reactive dyes form permanent covalent bonds with the cellulosic fibres to fixate to the fabric, resulting in bright and colourfast dyed fabric [12][13]

Cotton is primarily dyed using reactive dyes, which dominate industrial use because they form covalent bonds with cellulose, producing bright shades with excellent wash fastness. In contrast, direct dyes rely only on physical adsorption and weaker intermolecular forces, making them easier to remove but less durable. Given that reactive dyes account for 48.23% of the global cotton dyestuff market in 2024 (Mordor Intelligence, 2024), they represent the most commercially relevant and environmentally challenging dye class. For this reason, the project focuses on understanding and removing reactive dyes from cotton fibres, as their strong chemical fixation makes them significantly harder [17] to extract than direct dyes.

2.3 Hydrolysis Mechanism ?

A successful mechanism for the removal of dye from cotton fibres involves the swelling of cellulose. This can be done with a combination of water and a solvent able to dissolve cellulose. Swelling disrupts the physical interactions between dyes and polymers, reducing the strength of Van der Waals forces and reducing fibre density, hence allowing for the easier removal of dyes

[15]. This is the mechanism used by some ionic liquids, such as BMIMCl and EMIMAc, which can be applied to dye extraction due to their ability to disrupt hydrogen bonds in cellulose [14][13].

2.4 Previous Works and Ionic Liquid Selection

Previously the removal of reactive dyes has been investigated with varied temperatures, water contents and cycle times. However, instead of using ILs other solvents, most notably, N-methylmorpholine-N-oxide monohydrate (NMMO) [17], have been used to achieve this due to them sharing the ability to dissolve cellulose. In previous processes a pretreatment of NaOH or another alkaline solution is first used to cleave the covalent bonds between the reactive dye and the cotton fabric. This is then followed by using an NMMO-water solution to swell the cotton fibres and dissolve the dyes into solution. After this step is completed, a HCl solution is then added to neutralise the NaOH to NaCl and then the resulting liquid is distilled to remove water [16]. Similar experiments have also been conducted using other solvents like dimethyl sulfoxide (DMSO) for direct dye removal from cotton [17][18] and other alkali solutions to disrupt covalent bonds. Whilst both solvents, DMSO and NMMO are considered green solvents, the other reagents necessary can be harsh and destructive, for example NaOH and HCl which are both corrosive solutions. There are also disadvantages to the solvents themselves, as NMMO requires a high energy consumption for dissolution and is thermally unstable [19]. These processes also require multiple reagents and steps per cycle, as they include pre-treatments or acidifications for optimum activity, which is expensive, wasteful and time-consuming. They also require separation by acetone washing and vacuum distillation to separate water, dye and solvent before they can be used in another dyeing or stripping cycle which is energy and time intensive. In using ILs, many of these steps are not necessary as the only reagent used in the entire stripping cycle can be the chosen IL, which is the case for the DMBA[H₂SO₄] dye stripping cycle that is investigated in this paper. DMBA[H₂SO₄] instead works by hydrolysing the covalent bonds itself and allows the dye to dissolve into it, hence removing the need for alkali pretreatment. Furthermore, the dye doesn't need to be separated from it to dye other fabrics as it provides beneficial conditions for dyeing once diluted.

Other benefits of [DMBA][H₂SO₄] include economic aspects as it is a low-cost PIL with an estimated production cost of only \$1–3 per kg, and its IL properties listed (see 2.1 ILs) align closely with the requirements of cotton dye removal. These features make it both economically and operationally attractive for treating dyed cotton fibres, particularly in processes where solvent recovery, elevated temperatures, and reduced environmental impact are essential.

3. Methodology

3.1 Materials

Cotton and wool fabrics were purchased from Whaleys Fabrics (UK). Six different Procion MX dyes (Jacquard, USA), also known as dichlorothiazine dyes, were used: Reactive Yellow 86 (PMX 004), Reactive Orange (PMX 020), Reactive Red 5 (PMX 034), Reactive Red 11 (PMX 040), Reactive Blue 4 (PMX 072) and Reactive Grey (PMX 211).

3.2 Ionic Liquid Synthesis

>98.0% N-Butyldimethylamine (DMBA) (Tokyo Chemical Industries, Japan) and 72% HS_2O_4 (sulfuric acid) (Thermo Fisher Scientific, USA) were used to synthesise the IL. In an ice cooled round bottom flask, DMBA received dropwise equimolar sulfuric acid while being continuously stirred overnight. Deionised (DI) water was then added to obtain a 20 wt% water IL solution, as this composition was found to be optimal for promoting IL penetration into cotton while maintaining sufficient acidity to disrupt dye–fibre interactions (Firth et., al., 2024). A 1:1 acid base ratio was also reached and both water content and acidity parameters were measured using a V20 Volumetric Karl Fischer titration (Mettler Toledo, USA) and a G20S Compact Titrator (Mettler Toledo, USA) respectively.

3.3 Cotton Dyeing

For each dye colour, 20 g of cotton fabric was cut into rectangular pieces and prewashed with Synthrapol (Jacquard, USA) and DI water in a Pyrotec 4 dyeing machine (Roaches, UK) at 40 °C for 10 minutes to remove impurities and ensure uniform dyeing. After an initial cold-water rinse, dyeing was carried out following the Procion MX “Heavy Shade” protocol. Each tube received 0.65 g of dye, 500 g of DI water, and 18 g of NaCl to enhance dyebath exhaustion (Noah et al., 1986). The tubes were heated to 50 °C and agitated for 20 minutes, after which 2.53 g of soda ash (Na_2CO_3) was added to initiate fixation. Dyeing then continued for a further 50 minutes. The cotton samples were subsequently afterwashed with Synthrapol for 10 minutes, rinsed in cold water, and air-dried overnight.

3.4 Cotton hydrolysis, dye extraction and wash processes

Dye extraction was carried out in four main phases. In the first “virgin IL” phase, dye was removed from cotton using virgin IL. This was followed by three “recycled IL” phases, each using the IL recovered from the previous cycle to extract dye from new cotton samples.

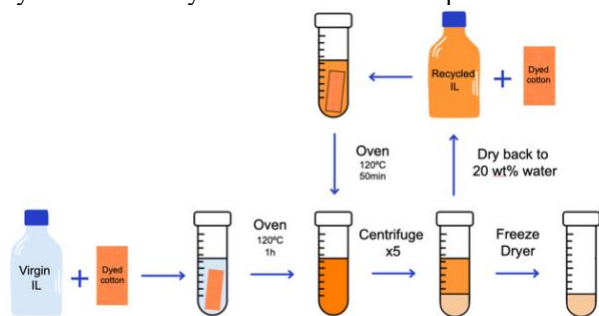


Figure 3.4: Schematic of the cotton hydrolysis and dye extraction process with IL recycling

3.4.1 Virgin IL dye extraction and Washing Process

Dye extraction was first carried out by mixing virgin IL with dyed cotton at a 4:1 IL-to-cotton ratio. Samples were prepared for each dye colour, including two Red 11 (PMX 040) samples, and each mixture was placed in a pressure tube and heated in a Heratherm oven (Thermo Fisher Scientific, USA) at 120 °C for 50 minutes to promote dye transfer from the cotton into the IL. After heating, the resulting hydrolysed cotton and coloured IL were mashed to ensure breakdown of fabric. An equal amount of water to IL was added to the mixture, which was then separated using an ST Plus Sorvall centrifuge (Thermo Fisher Scientific, USA) at 3600 rpm for 20 minutes resulting in a cotton pellet and a supernatant containing DI water, the IL and dye. The cotton pellet was then washed four more times with DI water at a 4:1 water-to-cotton ratio and centrifuged through Washes 2–5 for IL removal.

3.4.2 Conductivity Test

A conductivity test was performed prior to the recycled IL dye extraction process to determine the number of washes needed to separate the IL from the cotton. To avoid losing coloured samples, the measurement was instead carried out on an equivalent uncoloured cotton–IL system. Using a SevenExcellence S400 conductivity meter (Mettler Toledo, USA), each supernatant at multiple washes cycles was diluted with DI water to fall within the calibrated IL conductivity range (0–2 mS) and then measured. The IL concentration (wt%) in each wash was obtained from a previously established calibration curve, allowing to track IL migration from the cotton into the wash water across cycles and ultimately determine the number of washes required for the experiment to ensure distinct separation of IL and cotton.

$$IL \text{ wt\%} = a\kappa - b \quad (\text{eq. 1})$$

Where $IL \text{ wt\%}$ represents the weight fraction of IL in the supernatant, κ the conductivity in $\mu\text{S}\cdot\text{cm}^{-1}$ and the calibration curve coefficients $a = 4.731 \times 10^{-7}$ and $b = 4.273 \times 10^{-6}$.

3.4.3 Recycled IL dye extraction phases

Following the first “virgin IL” dye extraction, all wash fractions (Washes 1–5) were combined for each colour and dried back to 20% water content by using a R-300 Rotavapor (Buchi, Switzerland) for around 30 min at 80°C and 135 mbar, followed by Karl Fischer and acidity tests to confirm the recovered liquid matched the properties of the virgin IL. The next three recycle phases would repeat the same dye extraction process (see 3.4.1 Virgin IL dye extraction and Washing Process) except for the virgin IL. The recycled IL from the previous phase would now be employed as it was aimed to study

the recyclability of IL as well as its impact on dye reusability.

3.4.4 Freeze Drying

The final cotton pellets obtained from the four experimental phases were freeze-dried prior to further analysis. Freeze-drying is a crucial step because removing water without applying heat preserves the structural integrity of the cellulose and prevents thermal degradation or further hydrolysis from leftover IL. Producing a dry, stable cellulose powder is also important for downstream processing, as powdered cellulose can be reconstituted or blended to form new materials such as polycotton fibres (Costa et al., 2022). Wet cotton pellets were first snap-frozen in liquid nitrogen to ensure rapid solidification and prevent crystallisation damage. The frozen samples were then placed in a freeze-drying vacuum system (Labconco, USA) and dried for a minimum of two days at pressures below 0.1 bar. Under these conditions, water is removed by sublimation ensuring that subsequent analyses including DP measurements, yield calculations, and fibre characterisation are not influenced by residual moisture.

3.4.5 Intrinsic viscosity measurements

With the freeze-dried product obtained, an assessment of the cotton fibre integrity post dye extraction was done. Cotton pulp viscosity was measured by dissolving 140mg of freeze-dried cotton in 25 mL of 1.00 mol/l cupri-diethylenediamine (CED) solution (Aldrich Chemistry, USA) and 25 mL of DI water, along with copper wire pieces, in a squeeze bottle. The mixture was shaken in an Innova 42 Incubator Shaker (New Brunswick, USA) for 30 minutes to ensure complete dissolution. Viscosity was then determined using a pulp viscometer (United Scientific, USA) equipped with a water jacket connected to a constant 25 °C recirculating water bath (Huber Kältemaschinenbau, Germany). The efflux time of a fixed volume of the pulp solution was recorded 5 times and averaged to calculate viscosity using the following equation.

$$\eta_{rel} = ht_n \quad (\text{eq. 2})$$

Where η_{rel} is the viscosity ratio, h the viscometer constant equivalent to 0.0925 and t_n the efflux time [s]. A table extracted from ASTM D1795-13(2021) Standard Test Method [22], was then used following the obtained viscosity ratio to find $[\eta]c$ in which $[\eta]$ is the limiting viscosity number [ml/g] and c the concentration of the sample equal to 28 mg/ml. This finally allows to find $[\eta]$ itself using the equation:

$$[\eta] = \frac{[\eta]c}{c} \quad (\text{eq. 3})$$

Ultimately, the degree of polymerisation (DP) can be calculated using the following:

$$DP = \left(\frac{[\eta]}{Q'} \right)^{\frac{1}{\alpha}} \quad (\text{eq. 4})$$

Where the constants $Q' = 0.42$ and $\alpha = 1$ for DP<950 (Marx-Firgini 1978, Gruber 1981).

By finding the DP for some of the freeze-dried cotton samples, an assessment on the impact of IL recycles and its effects on the degradability of cotton could be done.

3.5 Wool Dyeing

Following the dye extraction's virgin IL and three recycling steps, each colours' supernatant from Recycle 3 was used to dye wool samples after dilution to 90% water content. Prior to this, one of the Red 11 (PMX 040) supernatant from the virgin IL phase was applied to wool at different water contents (50–90%) to identify the optimal dilution for dye uptake. For wool dyeing, the same protocol as cotton dyeing was followed, except with a higher temperature of dyeing, at 100°C (see 3.3 Cotton Dyeing).

3.6 Dye tests

3.6.1 UV-Vis Test

Each cycle's freeze-dried cotton powder and supernatant was tested using UV-2600 UV-Vis spectrophotometer (Shimadzu, Japan) in a wavelength range of 350nm-750nm (visible range) with a 2 nm slit width to measure the colour strength and dye concentration change from the reflectance and the absorption peaks for solid and liquid samples respectively. A baseline spectrum was first recorded using an empty holder to correct background signals. For liquid samples, although the principles of absorption rely on the Beer-Lambert law, no calibration curve was done to determine constants due to limited access to reactive dyes. Therefore, a percentage change approach was selected to calculate the evolution of colour and chromogenic structures of the dyes in liquid samples. As for the solid samples, the reflectance peaks of the different recycle powders were compared to the one produced from the virgin IL cycle using the spectra. The Kubelka-Munk function was also used to determine the colour strength of each sample.

$$K/S = \frac{(1-R)^2}{2R} \quad (\text{eq. 5})$$

Where K is the absorption coefficient of light, S the scattering coefficient of light and R the surface reflectance identified on the spectra. This enabled the calculation of the percentage decolourisation between two solid samples using the equation below:

$$\% \text{decolourisation} = \frac{K/S_{\text{recycle}} - K/S_{\text{initial}}}{K/S_{\text{initial}}}$$

(eq. 6)

3.6.2 Dye Degradation Test

Additional thermal stability tests were performed by heating the first Virgin IL phase's wash in the oven for an additional four hours for only one colour Red 11, with samples taken at one-hour intervals. In total, samples were heated at 120°C from 1 to 5 hours (Virgin IL phase already heated for 1 hour in the oven at 90°C). Unlike the dye extraction step, a lower temperature of 90°C was chosen to avoid hydrolysis with potential cotton remaining in the supernatant. This way, only the effect of heat on dye and IL could be studied. All relevant dye concentrations and changes were quantified using UV–Vis spectroscopy to study the effect of heat on dye degradation within the IL.

4. Results and Discussion

4.1 Ionic liquid properties

4.1.1 Conductivity

Conductivity measurements of the five supernatants obtained after each centrifugation step showed that the IL was progressively removed from the cotton with each wash. However, complete removal was not achieved even after five washing cycles as the IL concentration in Wash 5 reached 0.8 wt%, indicating that a small but non-negligible fraction of IL stayed within the cellulose fibres. Such residue influences next analyses such as DP measurements and contributes to incremental IL losses across recycling cycles. While additional washing would be required to achieve full IL recovery or complete fibre purification, this must be balanced against higher water consumption

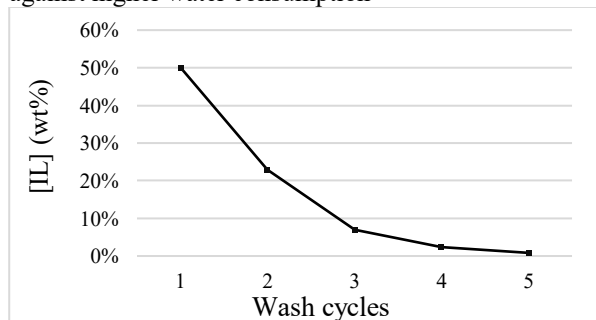


Figure 4.1: Evolution of the concentration of IL (wt%) in the supernatants of wash 1-5 for virgin cotton.

4.1.2 Acid-Base testing

Across all dyes, the acid:base ratio increased between the virgin IL and the third IL recycle. Increases ranged from +3.57% (Red 11) up to +6.93% (Orange), which can be explained by partial DMBA loss or chemical transformation during processing. Indeed, it is believed that because DMBA has a boiling point of 106 °C, prolonged heating and evaporation steps when processing the supernatant into the rotary evaporator at 80°C and a vacuum of 135 mbar had the potential to volatilise small amounts of the amine component. Loss of DMBA would shift the to a more acidic IL, increasing the acid:base ratio. Such acidity increases may influence

IL performance in subsequent cycles, including dye-stripping efficiency, fibre swelling behaviour, and susceptibility to dye degradation.

	Virgin IL	IL Recycle 3	%change
Red 11	1.017:1	1.053:1	3.570%
Orange	1.017:1	1.087:1	6.928%
Grey	1.017:1	1.072:1	5.435%
Red 5	1.017:1	1.063:1	4.515%
Blue 4	1.017:1	1.054:1	3.643%
Yellow 86	1.017:1	1.060:1	4.293%

Table 4.1: Acid:base ratio evolution from the first to the last cycle for all colours.

4.2 Dye Stripping in Cotton Fabric

4.2.1 Effectiveness of decolourisation

Each cycle produced dye stripped powder from the original dyed cotton fabric. The powder of each cycle was analysed with the UV-vis and compared to the cotton powder produced from the virgin IL cycle using the spectra and Kubelka-Munk equation to determine the % decolourisation of each cycle and assess if there had been any changes in chromogenic structure. [A change in colour was observed in the third recycle, evidenced by a significant change in peak wavelength for some colours. The colour most affected was blue which had a shift of approximately 45nm.]This was visible as the dyed cotton and the IL produced has a slightly greenish tinge. This change in chromogenic structure could be because of dye degradation due to the long times spent heating the ionic liquid and dye at high temperatures. This was further investigated by doing a dye degradation test as explained in section 3.6.2. Whilst some colours had no apparent chromogenic change from looking at spectra, to the naked eye there were slight changes, for example in grey which became more purplish with each cycle. For all colours there was a slight decrease in reflectance with each cycle, showing more strongly coloured cotton with each cycle. This could be because of inadequate washing of coloured IL in the cotton powder, making the solid appear to have more remaining dye than was actually in the fibres.

The Kubelka-Munk equation, as shown in Eq. 5 was used to calculate K/S values for each powder. The % decolourisation was then calculated by using Eq. 6 where the virgin IL powder of each respective colour was used as K/S initial. It was assumed decolourisation would be consistent since the same dyed fabric is used and the recycled washes should maintain the same properties, however this does not take dye degradation into account. The results of this did not follow the expectation that % decolourisation would be almost equal for each step, as the ILs potential to remove colour should be unaffected by the amount of dye present in the liquid. The wash is insufficient, which result in some loose dye being attached on the cotton powder surface. There was instead a high amount of variation

between cycles, however, there didn't seem to be a clear observed trend. The IL-dye solution colour that most effectively removed dye was orange which had an average decolourisation of 19.8% per cycle. The worst performer was grey which had only 1.7% decolourisation. This difference could be due to the molecular structure of the dye. There were also some cycles which appeared to be more saturated than the virgin IL powder. This could be due to fluctuations in colour of the original fabric being decoloured, since the values with negative decolourisations were small.

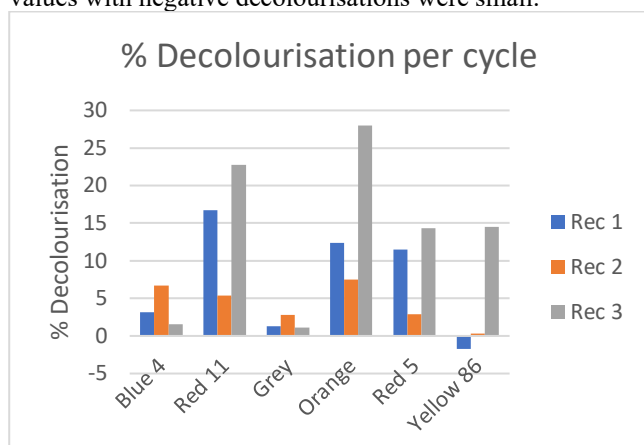


Figure 4.2: Decolourisation of each colour per cycle

4.2.2 Dye degradation

The results of the dye degradation test (see 3.6.2 Dye Degradation Test) show a range of changes in peak wavelengths.

$$\text{Peak \%change} = \frac{\lambda_{n \text{ hrs}} - \lambda_{1 \text{ hr}}}{\lambda_{1 \text{ hr}}} \quad (\text{eq. 7})$$

Where $\lambda_{1 \text{ hr}}$ is the absorption peak wavelength at 1hr and $\lambda_{n \text{ hrs}}$ the one at n hours with $n = 2, 3, 4$ or 5.

%	2h	3h	4h	5h
Red 11	-1.460	-1.460	-1.460	-1.642
Red 5	0.000	0.000	0.000	0.186
Yellow 86	0.000	0.000	0.000	0.000
Orange	1.414	0.808	1.212	1.212
Grey	5.703	2.281	3.422	4.563
Blue 4	0.669	-1.338	-2.007	-2.508

Table 4.2: Percentage change of each colour across 2-5 hours compared to the 1hour virgin IL supernatant solution.

From the table, percentage changes calculated using equation 7 are shown comparing each colour's evolution with their respective 1 hour absorbance peak wavelength. Blue 4 and Grey exhibit the highest change in absorbance peak wavelength, showcasing dye degradation and potential hydrolysis with the IL explaining the shift in absorbance peak wavelengths. On

another hand, the Red, Yellow and Orange show little to no change, where negligible dye degradation occurred. This difference is caused by the molecular structure of the reactive dyes. For instance by looking at Figure X, Reactive Blue 4 is an anthraquinone, in which C=O bonds makes the quinone ring reactive the more acidic IL, degrading the colour and intensities. In contrast, Reactive Yellow 86 has a single azo molecular structure with (–N=N–) bonds generally stable in strongly acidic conditions

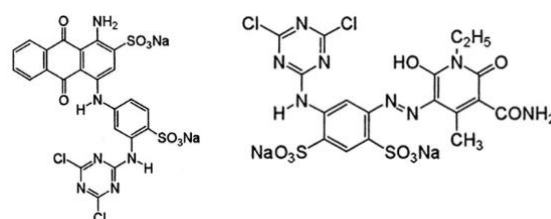


Figure 4.2.1: Molecular structure of Reactive Blue 4 (left) and Reactive Yellow 86 (right)

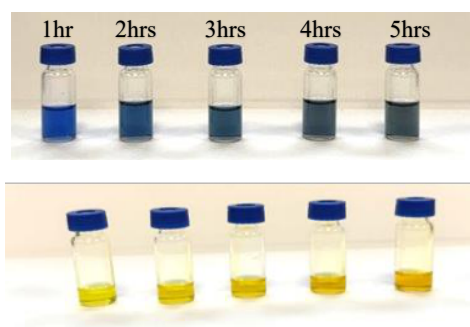


Figure 4.2.2: Evolution of Reactive Blue 4 and Reactive Yellow 86 IL supernatants overtime.

From the pictures above, although the peak shift stayed below a small 2.508% range in the UV-Vis spectra, the colour change is noticeable to the naked eye. This showcases the limitations of the range used in this study, where larger UV-Vis spectrophotometry range would allow more accurate reflection of colour change. [21]

4.3 Wool

4.3.1 Effect of Water Content

Five samples of wool fabric were first dyed using variations of the virgin IL wash of the Reactive Red 11 fabric each made up to a water content of 50-90%. This was done to determine the optimal water content wool dyeing since water content impacts the acidity. It was not necessary to separate the dye from the IL for this since the IL was being significantly diluted so wouldn't cause dye removal. Considering wool's need for an acidic dyebath, the acidic IL aids dye conditions since the functional groups of the reactive dyes are hydrolysed during dye stripping so now behave as acid dyes. The acidic environment allows the amino groups in wool to be protonated, so they are able to ionic bond with the dyes.

An outlier occurred at 70% water content, which exhibited a pale coloured fabric and a much higher

reflectance than the other samples. It is suspected this happened due to residual acetone in the metal flasks the wool was dyed in from previous users. Otherwise, it could be seen that fabrics dyed with ILs of higher water content had lower reflectance, demonstrating that they had a higher colour uptake so better absorbed the dyes. Therefore, IL with 90% water was determined to be the most effective. This is supported by the concept that higher water content in DMBA[H₂SO₄] provides more acidic conditions which would further assist acid dyeing [19].

4.3.2 Wool Dyeing in DMBA[H₂SO₄]

Following from the results of section 4.3.1. It was decided that wool samples would be dyed with 90% water content DMBA[H₂SO₄] dye solution. The dyed wool samples were then analysed by the UV-vis spectrophotometer and then analysed using their resulting spectra and the Kubelka-Munk equation (Eq. 5). The resulting dyed wool fabrics displayed very different colours to the previously water dyed cotton, verified by shifted peaks when measured in the UV-vis spectrophotometer. The shape of the graphs was also very altered, as they were flatter and had indistinct peaks compared to previous dyed fabrics. This could be due to residual soap in the tubes interfering with the dye or possible but unlikely extreme degradation of dye during the dyeing process. The UV-vis spectrum of soap supports this idea as it shows a flat graph with no visible peaks within the 350-750nm wavelength range. The colours most affected by this were pink and blue which had very noticeable colour changes, while some colours, like orange and yellow, had little peak wavelength change indicating little change in chromogenic structure between the dyed cotton to the dyed wool. Nonetheless, there seems to be good dye uptake as most the water dyed cotton colours achieved a similar K/S value to fabric for their respective colour. This could be due to dye structure.

Building on these observations, the dye uptake on wool was not uniform across all colours. Analysis of the supernatant from before and after the wool dyeing process was done showing a clear difference: grey and orange showed the highest uptake according to Table 4.3.2. This uneven behaviour suggests that specific dye wool interactions impacted the dye uptake. According to Figure 4.3.2, the absorption peaks of Blue 4, Red 11 and Red 5 were hardly distinguishable, therefore the dye uptake was not calculated. This is caused by the supernatant post dyeing becoming clear after dye transfer which made the peaks become wider and flatter due to the lighter colour.

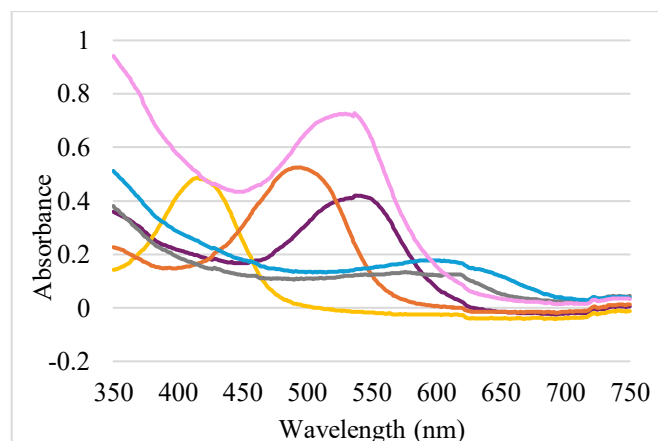


Figure 4.3.2: UV-Vis absorption spectra of IL Recycle 3 supernatant before wool dyeing.

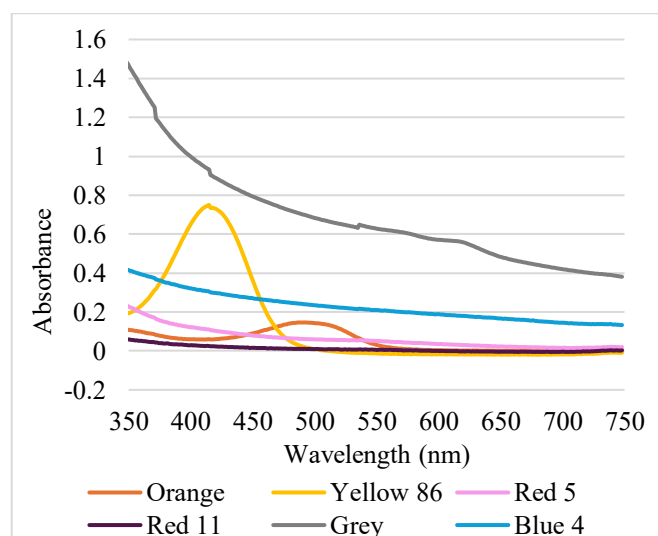


Figure 4.3.3: UV-Vis absorption spectra of IL Recycle 3 supernatant after wool dyeing.

%change of peaks before and after wash dye

$$\%change = \frac{\lambda_{after\ dye} - \lambda_{before\ dye}}{\lambda_{before\ dye}} \quad (\text{eq. 8})$$

	Before (nm)	After (nm)	%Change
Grey	576	621	7.81%
Yellow 86	416	405	-2.64%
Orange	494	480	-2.83%

Table 4.3.2: Percentage change of absorption peak wavelengths of the IL Recycle 3 supernatant before and after wool dyeing.

The absorption peaks intensity also changed, and calculating the difference represents that a dye uptake of 88% was reached for Grey for the process.



Figure 4.3.4: Comparison of water dyed cotton (left) and IL recycle 3 dyed wool (right).

4.4 Cotton Analysis

4.4.1 Cotton Yield

Cotton yield averaged $91.8\% \pm 0.1\%$ across all dyes and IL cycles, which is expected because several steps in the process inevitably lead to small mass losses. During IL treatment and heating, a portion of the cellulose undergoes partial hydrolysis, producing fine particles that may not fully sediment during centrifugation and are therefore lost in the supernatant. Additional losses can occur during washing, transfer between tubes, and freeze-drying, where very small fibres or lint adhere to container walls or remain suspended in wash water. Specific drops in yield (e.g., Recycle 2 for Red 11 and Recycle 3 for Red 5) were attributed to handling errors. Finally, these measurements assume that the original dyed cotton feedstock and the freeze dried cotton samples are fully dry. In reality, some water can still be present affecting the accuracy of the cotton yield.

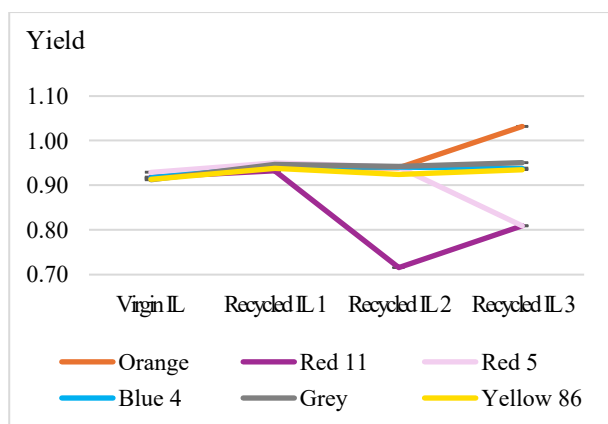


Figure 4.4.1: Cotton yield evolution across IL recycles for each colour

4.4.2 Cotton Integrity



Figure 4.4.2: Microscope picture of Orange and Pink Recycle 3 powdered cotton fibres.

From following the methodology in section 3.4.5 testing the viscometry of the Red 5 and orange powdered cotton samples in a viscometer a mean time was obtained per cycle for each respective colour. Both colours showed a general decrease in the degree of polymerisation with each cycle showing an increase in cotton fibre degradation with each cycle. This shows that the IL becomes more degrading with each recycle and has a larger effect on cotton fibre integrity with time and heat. This could be due to the acid-base ratio not being 1:1 in subsequent cycles, causing degradation. Both the pink and orange virgin IL cycle had a degree of polymerisation of 250 ± 2.866013 .

Orange degraded slightly more than pink, ending up with a DP of 221 in recycle 3 giving a 11.4% decrease in DP compared to 228 for pink in recycle 3 giving an 7.7% drop in DP. There were some outliers where degree of polymerisation increased between cycles. This was likely due to the variance in the manual mashing process when the hydrolysed cotton is being powdered.

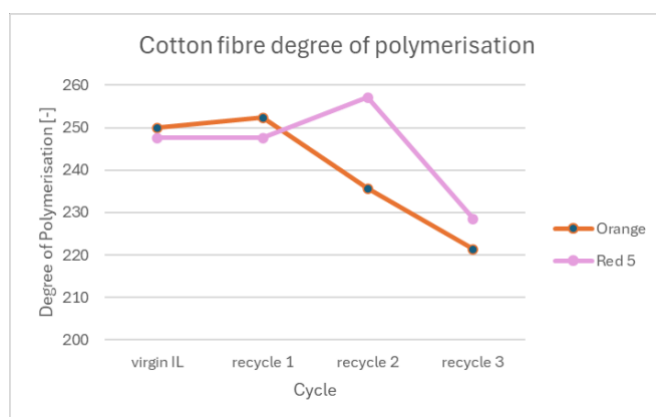


Figure 4.4.2: degree of polymerisation for reactive red 5 and orange dyed cotton fibres

Conclusion and Future Work

In conclusion, in the process of reactive dye stripping from cotton fibres, an overall cotton yield of $91.8\% \pm$ was achieved. This loss is too high for industrial practices which allow only a loss less than 1%. To reduce this, stronger centrifugation should be employed to ensure proper sedimentation of particles as well as

thorough washing to prevent fibres remaining on apparatus, and the avoidance of unnecessary transfers between containers. There is a slight degradation of cotton as the number of cycles increases, which is hypothesised to be due to the increase in acidity of the ionic liquid due to DMBA evaporation during the drying of the supernatant. To combat this, further investigation should be done to test if a decreased acid-base ratio would be more effective in avoiding this. In terms of dye removal, the efficiency of DMBA[HSO₄] in decolourising cotton fabric varied largely depending on the colour of the dye. This was probably caused due to differing chemical structures, causing some dyes to attach more strongly to the cotton fibres, hence making them harder to remove. With the stripped dye, it was found that wool dyeing with the stripped dyes in 90 wt% water DMBA[HSO₄] as possible with similar dye uptake to cotton fibres as they achieved similar K/S values. The acidic environment provided by the water content aided dye uptake, but the optimum acidity and corresponding water content should be determined using the Hammett acidity test. However, a big colour change from cotton to wool fibre was observed, which was demonstrated by very differently shaped UV-vis spectra. This could be due to several factors such as, residual soap interference with the dye, dye degradation in the dyeing process, or reaction with the wool fibres.

In future, to further explore this research area, more experimentation can be done in using the stripped dyes to dye other fabrics, such as nylon and silk which should also be theoretically able to be dyed with acid dyes. In addition to this, the concept of polycotton degradation and separation using DMBA[HSO₄] could be explored. This would involve the addition of disperse dyes and polyester fabric into the experiments to investigate how this complication would affect the efficiency of the process. Finally, more work can be done to optimise the recovery of the ionic liquid, as this would make the process more sustainable and more economical.

Acknowledgement

The authors gratefully acknowledge the support and supervision of Dr Xuantong Sun during the execution of this study. Her technical guidance and constructive input were essential to ensuring the quality of the research.

References

1. WasteManaged. Fashion waste facts and statistics [Internet]. 2023 [cited 2025 Jan XX]. Available from: <https://www.wastemanaged.co.uk/our-news/retail/fashion-waste-facts-and-statistics/>
2. United Nations Environment Programme (UNEP). Unsustainable fashion and textiles focus on International Day of Zero Waste [Internet]. 2023 [cited 2025 Jan XX]. Available from: <https://www.unep.org/news-and-stories/press-release/unsustainable-fashion-and-textiles-focus-international-day-zero>
3. Earth.org. Fast fashion's detrimental effect on the environment [Internet]. 2023 [cited 2025 Jan XX]. Available from: <https://earth.org/fast-fashions-detrimental-effect-on-the-environment/>
4. Alchemie Technology. Toxic soup: dyeing chemicals [Internet]. 2022 [cited 2025 Jan XX]. Available from: <https://www.alchemietechnology.com/news/toxic-soup-dyeing-chemicals/>
5. Yaseen DA, Scholz M. Textile dye wastewater characteristics and constituents of synthetic effluents: a critical review. *Water*. 2019;11(9):1985. Available from: <https://pubs.ncbi.nlm.nih.gov/articles/PMC7490536/>
6. Islam MM, Mahmud K, Faruk MO, Billah MS, Islam A. Textile dyeing industries in Bangladesh: pollution management and treatment of wastewater. *J Textile Sci Eng* [Internet]. 2011 [cited 2025 Jan XX];1:101–6. Available from: <https://www.scirp.org/journal/paperinformation?paperid=17027>
7. RSC Publishing. Article D3NR06126D [Internet]. 2024 [cited 2025 Jan XX]. Available from: <https://pubs.rsc.org/en/content/articlehtml/2024/nr/d3nr06126d>
8. Textile Exchange. Cotton overview [Internet]. 2024 [cited 2025 Jan XX]. Available from: <https://textileexchange.org/cotton/>
9. Textile Exchange. 2025 Preferred Fiber & Materials Market Report [Internet]. 2025 [cited 2025 Jan XX]. Available from: <https://textileexchange.org/knowledge-center/reports/materials-market-report-2025/>
10. Feng L, Chen Z. Research progress on dissolution and functional modification of cellulose in ionic liquids. *J Mol Liq*. 2008;142:1–5.
11. Baaqel H, Diaz I, Tulus V, Chachuat B, Guillen-Gosalbez G, Hallett JP. Role of life-cycle externalities in the valuation of protic ionic liquids: a case study in biomass pretreatment solvents. *Green Chem*. 2020;22(10):3132–40.
12. Alfa Chemistry. Scientific overview of reactive dyes and their industrial applications [Internet]. 2023 [cited 2025 Jan XX]. Available from: <https://www.alfa-chemistry.com/dyes/resources/scientific-overview-of-reactive-dyes-and-their-industrial-applications.html>

13. Encyclopaedia Britannica. Dyeing techniques [Internet]. 2023 [cited 2025 Jan XX]. Available from: <https://www.britannica.com/technology/dye/Dyeing-techniques>
14. Wang S, Wu P, et al. Profitable and environmentally friendly dyeing strategies. *Chem Eng J* [Internet]. 2024 [cited 2025 Jan XX]. Available from: <https://www.sciencedirect.com/science/article/pii/S1385894723048817>
15. Sustainable cotton decolourisation paper. *J Cleaner Prod* [Internet]. 2024 [cited 2025 Jan XX]. Available from: <https://www.sciencedirect.com/science/article/pii/S2352186424004164>
16. Complete separation of dye/cotton systems. *Chem Eng J* [Internet]. 2021 [cited 2025 Jan XX]. Available from: <https://www.sciencedirect.com/science/article/pii/S138589472103151X>
17. Aracri E, Vidal T. Cellulose reactivity and dissolution. *Polym Degrad Stab* [Internet]. 2012 [cited 2025 Jan XX]. Available from: <https://www.sciencedirect.com/science/article/pii/S0032386112009123>
18. Sjöström E. Wood chemistry and cellulose structure. *Cellulose* [Internet]. 2007 [cited 2025 Jan XX];14:599–610. Available from: <https://link.springer.com/article/10.1007/s10570-007-9160-x>
19. Hallett JP, Welton T. Room-temperature ionic liquids dissolve cellulose. *Chem Rev*. 2011;111:3508–76.
20. Agarwal UP, Ralph SA, Baez C. Hydrolysis of cellulose in ionic liquids. *Green Chem* [Internet]. 2013 [cited 2025 Jan XX]. Available from: <https://pubs.rsc.org/en/content/articlelanding/2013/gc/c3gc40136g>
21. Starczak R, Was-Gubala J. Colour as a basic feature in comparative studies of fibre microtraces. *Issues Forensic Sci*. 2015;290:104–8. doi:10.34836/pk.2015.290.5.
22. ASTM International. ASTM D1795-13(2021) Standard Test Method for Intrinsic Viscosity of Cellulose [Internet]. West Conshohocken, PA: ASTM International; 2021 [cited 2025 Jan XX]. Available from: <https://www.astm.org/d1795-13r21.html>. [ASTM International | ASTM](#)

Exploring the effect of amphiphile morphology of APGs on interfacial tension

Abid Sarker and Mohammed Nadim Miah

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Sugar-derived surfactants such as alkyl polyglucosides (APGs) are rapidly gaining industrial relevance due to tightening environmental regulations and the growing demand for sustainable solutions to legacy petrochemical surfactants. Despite their commercial potential, the design of APGs remain largely empirical due to the limited understanding of how subtle changes in molecular morphology impact interfacial performance. This study leverages atomistic molecular dynamics simulations to examine the effects of branching architecture on the interfacial behaviour of APG8 and its variants at a water-oil interface. Through a hybrid forcefield model, the interfacial tension was quantified from a local pressure anisotropy. The critical micelle concentrations, identified from the density peaks at the interfaces, are used to translate surface coverage results into corresponding bulk concentrations through the Gibbs thermodynamic framework. This provided insights into how amphiphile morphology governs adsorption efficiency. Clear relationships are observed across the different surfactant groups tested. Symmetric double-branched variants of APG8 produced enhanced reductions in interfacial tension relative to the linear option, whereas the performance of triple-branched systems is strongly influenced by both branching geometry and position of the branch points along the tail. These findings lay a foundation for the rational development of future sustainable amphiphiles.

1 Introduction

Surfactants are central to many industrial and consumer processes due to their ability to modify interactions between immiscible fluids. This can be in detergents, cosmetics, pharmaceuticals or energy technologies. The capacity of surfactants being able to reduce the interfacial tension and promote mixing, spreading or stabilisation of immiscible phases underpins products used daily by billions of people. The global demand for surfactants reflects this ubiquity, with the market exceeding \$ 45-50 billion in 2023 and is projected to surpass \$ 65 billion by 2030, driven by population growth, rising hygiene expectations and expansions of home and personal care sectors in emerging markets [1][2]. Detergents represent the largest share of surfactant consumption, with cleaning products being the primary application. Traditional petrochemical derived surfactants are facing growing scrutiny. Prominent surfactant classes such as linear alkylbenzene sulfonates (LAS) and nonylphenol ethoxylates (NPEs) pose many environmental and toxicological concerns. These include poor biodegradability, formation of harmful degradation products and threat to aquatic environments, which are now becoming recognised by global regulatory frameworks [3]. In addition, consumers are starting to prioritise products that are derived from renewable feedstocks and have low environmental impact. Industries are shifting towards more sustainable and environmentally responsible resources, with a spark in interest in developing surfactants that can match or even exceed the performance of the traditional fossil fuel-based options.

Among the most promising candidates in this transition are alkyl polyglucosides (APGs), a family of non-ionic and sugar-derived surfactants.

They are synthesised from glucose and fatty alcohols. APGs provide the renewable sourcing as well as low toxicity and excellent compatibility, making them attractive alternatives in many application areas. Recent analyses predict that the bio-surfactants industry, propelled largely by APGs, will grow at over 10% annually through 2030, outpacing conventional surfactants [4]. This emphasises the commercial potential and the importance of accelerating the understanding of how APG structures behave. However, finding the optimal profile remains the biggest challenge, as APGs encompass a broad spectrum of possible structural configurations. This can include morphological variations in the length of the hydrophobic tail, how the sugar head connects to the tail, or the presence of branching; all of which influence how effectively the molecule performs at an interface. Even subtle structural changes can strongly influence key interfacial properties such as surface tension, adsorption efficiency and the concentration at which micelles start to form. Investigating these effects through experiments would require an extensive amount of testing, resulting in a slow and expensive process.

Advancements in modern computational tools have made the process of analysing these intricate details very accessible. In particular, the rise of molecular dynamics (MD) has provided valuable insights into how APGs behave at certain interfaces. By observing how individual molecules arrange themselves and interact with the neighbouring phases, MD helps identify the structural factors that make certain APGs more effective than others. While previous studies have focused on specific APGs, notably APG12, there remains a need to explore how broader structural modifications influence behaviour at more complex

interfaces, particularly in shorter-tail surfactants like APG8, an area that remains relatively underexplored.

2 Background

Surfactants, short for surface active agents, are amphiphilic molecules containing a hydrophilic functional group, which has an affinity for water, and a hydrophobic alkyl chain, which repels water. They localise at interfaces, such as water-air or water-oil, reducing surface tension and enable processes like emulsification, foaming, wetting and detergency to occur. They are classified depending on the nature of the hydrophilic group. There are 4 main types: non-ionic, anionic, cationic and zwitterionic. The hydrophobic tail group of most surfactants is made up of hydrocarbons. Each type is used for different purposes such as detergents, cosmetics, pesticides, drug delivery and so on. The performance of a surfactant is determined by its ability to efficiently pack at interfaces and minimise surface energy.

Surface tension is a key parameter in characterising a surfactant's performance. It represents the free energy per unit area required to create an interface between two phases. The addition of surfactants decreases the surface tension through the replacement of high energy solvent-solvent interactions with more stable surfactant-solvent interactions. The rate at which these effects occur depends on the surfactant's structure, concentration and orientation at the interface. This occurs until a saturation point, known as the critical micelle concentration (CMC). Beyond this point, the energy required to insert more surfactants in will be greater than the energy needed for surfactant molecules to form micelles in the bulk solution. This is the concentration at which surfactants self-assemble into micelles through aggregation of its hydrophilic heads towards the aqueous phase, reducing the additional free energy. Micelles are normally spherical in shape, but they can also form in different shapes depending on conditions.

3 Objectives

The objective of this paper is to examine the effects of amphiphile morphology and its influence on surface tension reduction capabilities of APGs through atomistic MD simulations. It builds on from the methodologies and insights discussed in the thesis by Kamrul-Bahrin [5], which primarily focused on APG12 and its structural variations. In contrast, this study shifts focus on APG8, which was

the most effective linear-chain APG, and explores a more complex and industrially relevant water-oil interface that is still largely unexplored.

By exploring branching and newly proposed triple-branching APG8 structures, the aim is to assess how these morphological changes affect the packing behaviour of surfactants at the interface. By taking the standard APG8 and APG12 as reference systems, this work seeks to develop a clearer understanding of the factors that influence the performance of sugar-derived surfactants and generate insights to help guide the development of future sustainable amphiphiles.

4 Methodologies

4.1 Interfacial Tension

The interfacial tension, γ , between two phases can be obtained through the Hulshof integral, which evaluates the inhomogeneity of the pressure tensor. It involves looking at the difference between the normal and tangential terms, shown in equation (1), which respectively correspond to the macroscopic surface tension and the microscopic pressure variations across an interface.

$$\gamma = \int_{-\infty}^{+\infty} [P_N - P_T] dz \quad (1)$$

where P_N denotes the “normal pressure” and P_T the “tangential pressure”.

Interfacial tension arises from the differences in stresses and the Hulshof integral quantifies it. In simple terms, as one moves along the surface in the z -axis, the pressure tensor becomes uneven due to the rearrangement of molecules. The integral of this pressure anisotropy outputs the work required to increase the interfacial area. However, this concept relies on global pressure averages and hence is not applicable when looking at inhomogeneous systems such as water-oil, where spatial resolution is required to explore the physics of an interface. An alternative is the Irving-Kirkwood (IK) local formulation.

$$P_{kk}(z) = k_B T \rho(z) + \frac{1}{A} \left\langle \sum_i^{N-1} \sum_{j<1}^N \frac{1}{|z_i - z_j|} (f_{ij})_k (r_{ij})_k \right\rangle \quad (2)$$

with k_B the Boltzmann constant, T temperature, $\rho(z)$ the density along z , A the interfacial area, N the number of molecules and the double sum of the configurational contribution averaged over the

ensemble. Dividing the box into n slabs along z , one evaluates $P_{kk}(z)$ and obtains γ from:

$$\gamma = \int_{-\infty}^{+\infty} \left[\langle P_{zz}(z) \rangle - \frac{\langle P_{xx}(z) \rangle + \langle P_{yy}(z) \rangle}{2} \right] dz \quad (3)$$

This approach allows the evaluation of diagonal elements of the inhomogeneous pressure tensor (P_{xx} , P_{yy} , P_{zz}) in terms of the coordinate perpendicular to the interface. This is easily done through computational software when conducting MD simulations and is applied to all systems containing surfactants such as APG12, APG8 and branched APGs tested in this study. By obtaining these spatial components, the interfacial tension can be determined.

4.2 Density peaks

To determine the critical micelle concentration (CMC) from the MD simulations, an approach based on the maximum density of surfactant molecules across the interface, following the methodology discussed by Liu et al [6] and Jang et al [7] was employed. Although previous studies have proposed identifying the CMC by examining the interfacial thickness and the angle of the surfactant head, reports in the literature indicate that these methods can be inconsistent and difficult to analyse with certainty. Hence, the decision is to focus exclusively on the density peak criterion, which provides a more robust and reproducible indicator of micellisation point.

The density profile of the surfactant normal to the interface reflects how surfactant molecules organise with the interfacial region. As the surface coverage (Γ) increases, the peak of this density distribution, initially increases because surfactant molecules pack more efficiently at the interface. However, once the interface becomes saturated, corresponding to the CMC, further addition of surfactants causes the interface to expand, destabilise and undergo significant fluctuations and appear ‘wrinkled’. This leads to a decrease in the measured density peak despite the higher number of surfactant molecules. In a macroscopic system this transition corresponds to surfactant molecules leaving the interface to form micelles. In these simulations, the metastable overpacked interface indicates this behaviour. This behaviour can be visualised in a larger simulation system using coarse grain models. However, this is not within the scope

of this project. Finally, the CMC is identified as the coverage corresponding to the maximum of the density peak curve. This value agrees with the physical expectation that the interface can accommodate only a finite number of surfactant molecules before a metastable interface exists and excess surfactants preferentially migrate to the bulk.

4.3 Bulk Concentration

To quantify the interfacial behaviour of APG8 at the water/decane interface, the bulk surfactant concentration associated with each simulated surface coverage was obtained directly from equilibrium MD trajectories. Following the approach of Yoo et al. [8], this mapping establishes a connection between the microscopic interfacial organisation and the macroscopic surface-tension isotherm $\gamma(c)$, allowing direct comparison with experimental measurements and identification of the CMC. The equilibrium distribution of surfactant molecules between the interface and the bulk phase is determined by the requirement that their chemical potentials be equal:

$$\mu_s = \mu_B \quad (4)$$

Where μ_s is the interfacial chemical potential of adsorbed surfactant, and μ_b is the chemical potential of surfactants in the bulk. However, μ is only a function of concentration in isothermal conditions. Below the CMC the bulk solution is extremely dilute, and the ideal-solution expression for chemical potential applies:

$$\mu_B = \mu_B^{ideal} = \mu_B^0 + k_B T \ln(x_B) \quad (5)$$

Where μ_B^0 is the infinite-dilution reference chemical potential in the bulk, whilst the last term is the ideal mixing (entropy) contribution. x_B is the bulk mole fraction, and since the solution is dilute it can be related by

$$x_B = \frac{C}{C_w} \quad (6)$$

with C the molar concentration (mol/L) of the surfactant in the bulk solution and C_w the molar density of water.

The interfacial chemical potential of the surfactant molecules adsorbed at the interface is decomposed into ideal and excess contributions:

$$\mu_s = \mu_s^{ideal} + \mu_s^{excess}(\Gamma) = \mu_s^0 + k_b T \ln(\Gamma a) + \mu_s^{excess}(\Gamma) \quad (7)$$

where μ_S^0 is the infinite dilution standard state chemical potential, and a is a normalisation constant equivalent to the area per surfactant head but can be approximated as $\frac{1}{\Gamma_{CMC}}$. Γ (nm^{-1}) is the surface coverage of the interface.

Through substituting equation (5) and (7) into (4), the following equation is derived:

$$\ln(x_B) = \frac{\Delta\mu^0}{k_bT} + \ln(\Gamma a) + \frac{\mu_S^{excess}}{k_bT} \quad (8)$$

Where $\Delta\mu^0 = \mu_S^0 - \mu_B^0$.

Each APG8 derivative has a slightly different $\Delta\mu^0$ but can be approximated to the value of APG8. As past MD simulation results on APG12 branching [6], suggests that branching only slightly alters the value. Therefore, in this project, values for $\Delta\mu^0$ for the branched versions of APG8 have been approximated to the value of APG8 at the water/decane interface.

The μ_S^{excess} term accounts for the non-idealities from the interactions of adsorbed surfactant molecules. At $\Gamma = 0$, the value of μ_S^{excess} is 0 whilst adding APG molecules increasing this term. These can be found through integrating the Gibbs adsorption equation.

$$d\mu = \frac{1}{\Gamma} d\Pi \quad (9)$$

At isothermal conditions, one can integrate the following equation with a reference state value of $\mu_S^{excess} = 0$ at $\Gamma = 0$.

$$\mu_S^{excess}(\Gamma) = \int_0^\Gamma \frac{1}{\Gamma'} \frac{d\Pi}{d\Gamma'} d\Gamma' \quad (10)$$

The effect on the Gibbs free energy from adsorbing APG molecules onto the interface is described by the μ_S^{excess} term. This equation can be solved by using a 2-dimensional equation of state which describes the interfacial behavior of adsorbed surfactants up to the CMC point.

$$\frac{\Pi}{k_bT} = \Gamma \left(1 + \sum_{i=1} B_{i+1} \Gamma^i \right) \quad (11)$$

After using tension data to find the B parameters (truncating after the third term), the derivative of the equation of state with respect to Γ' can be substituted into equation (10) in order to solve it using measurable quantities from our MD

simulations. Therefore, all quantities of equation (8) can be found and a relation for surface coverage to bulk concentration can be found.

4.4 Simulation Details

To investigate the effects of morphological changes, the molecular dynamic systems were modelled through Medea, provided by Material Design Inc. This software provided a platform to build the required atomistic models and generate the initial input files. These files were then exported to the High Performance Computing (HPC) systems at Imperial College London, where simulations were executed manually. All simulations were carried out using the LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) molecular dynamics engine [9], which serves as a computational backend for Medea.

Throughout this report, n-dodecyl β -glucoside, or APG12, consisting of a single sugar head with n-dodecane as hydrophobic tail, is treated as the reference structure. Building on the findings of Kamrul-Bahrin [5], which discovered that APG8 produced the lowest surface tension among the tail length variants and that branched structures further exhibited lower surface tensions due to improved interfacial packing; this report focuses on examining APG8 and a series of its branched derivatives as show in Figure 1. This includes two double-branched structures, APG7,1 and APG5,3, three triple-branched structures, APG6,1,1, APG5,2,1 and APG4,2,2 and one distal triple-branched structure, APG4,2,2(3-3). Due to time constraints within this project, only one distal branched structure was investigated.

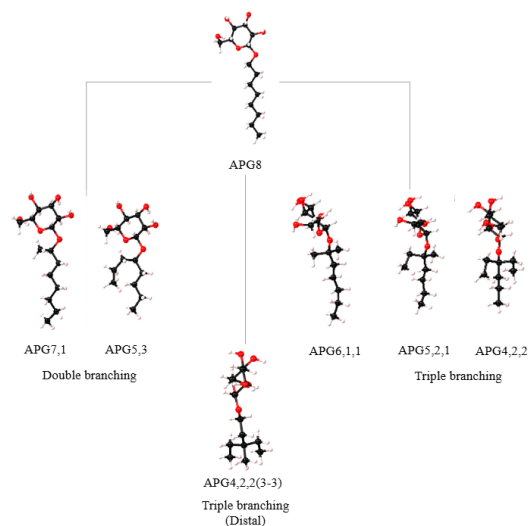


Figure 1. Representation of surfactants being tested

The initial simulation cell is an elongated orthorhombic box, where the dimension in the z direction is approximately three times that of the x

and y directions ($L_x = L_y$ and $L_z = 3L_x$). All three Cartesian coordinates have periodic boundary conditions applied to them. The long-range Coulombic interactions were calculated using the Particle-Particle/Particle-Mesh (PPPM) approach.

The simulations were run on an isothermal-isobaric (NPzAT) ensemble, maintained at a constant pressure in the z direction using a Nosé - Hoover barostat. The NPzAT stage was ran for 600 ps, with a timestep of 1 fs. This was then passed onto a canonical (NVT) ensemble, with a Nose-Hoover thermostat at constant temperature of 298.15 K, after equilibration. The constants for the barostat and thermostat were 1.0 and 0.2 ps for relaxation purposes.

Each system contained approximately 6000 water molecules, 800 decane molecules and a variable number of surfactants from 16 to 169 APG molecules corresponding to surface coverages from 0.444 to 2.8 molecules nm^{-2} . The simulation cell was formed by stacking layers in the order: decane-APG-water-APG-decane. This created two distinct water-decane interfaces. The simulations were carried out at $T = 298.15$ K for 2 ns using a 1 fs integration time step.

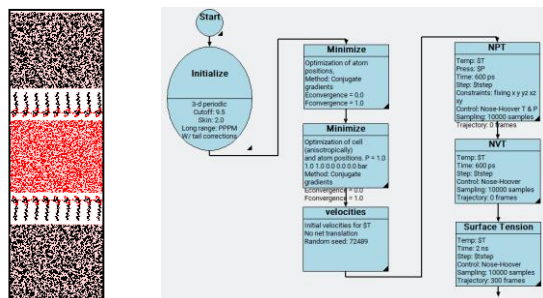


Figure 2. Simulation Box and LAMMPS process

4.5 Force fields

In this project, fully atomistic force fields were used to capture the molecular detail required to compare different APG variants. Although coarse-grained models for glycolipids, they need extensive fine tuning to obtain reliable interfacial tension, making it difficult to identify differences between structures. Thus, the simulations employed the atomistic force field, derived from the Polymer Compatible Force Field (PCFF) and the Polymer Consortia (PCFF+), provided within the Medea environment. This accurately predicted the liquid properties and molecular conformations.

When conducting initial test simulations, the standard PCFF+ water model produced inconsistent behaviour at the water-decane interfaces, with unstable surfactant layer and density fluctuations. To resolve this, the original PCFF (COMPASS) water model was used for water

molecules, whereas the decane and APG molecules was constructed using the PCFF+ parameters. This hybrid PCFF/PCFF+ model corrected the initial anomalies and provided stable behaviour for APG at the water-decane interface.

5 Results and Discussion

5.1 Interfacial Tension

To effectively evaluate how subtle changes in amphiphile morphology influence interfacial behaviour, it was necessary to quantify how each APG variant modifies interfacial tension as a function of surface coverage, Γ . This allowed for a visual and mechanistic understanding of how efficiently different structures pack at the interface, how early they saturate the surface and how strongly they disrupt the cohesive stresses at the water-oil boundary. This serves as the primary diagnostic for comparing molecular variants in small-scale MD systems, as mentioned by Cardenas et al [10].

Initially, it was important to analyse the reference systems, APG12 and APG8, to make sure they were reproducible. This was to verify that the workflow, simulation protocols and force-field selection were consistent relative to the results reported by Kamrul-Bahrin [5]. By establishing this agreement, it was confirmed that any subsequent discrepancies observed for branched and triple-branched APG8 variants could confidently be attributed to structural effects rather than changes in the methodological approach.

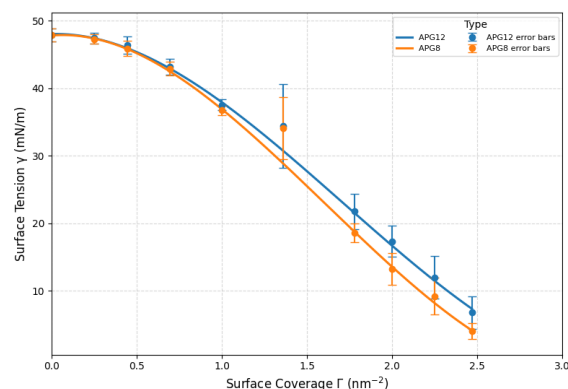


Figure 3. Interfacial tension against surface coverage plot for reference systems

Both APG12 and APG8 showed the expected monotonic decrease in γ as Γ increases, which is consistent with the classical surfactant adsorption theory. APG8 consistently yields lower surface tensions at equivalent surface coverages compared to APG12 due to higher surface activity

per molecule. This reflects how the shorter APG8 tail can pack more densely and with less steric restrictions, leading to more efficient saturation at the interface. Hence, this showcases that APG8 is clearly the strongest performer among linear chains and provides a baseline against which branched morphologies can be compared.

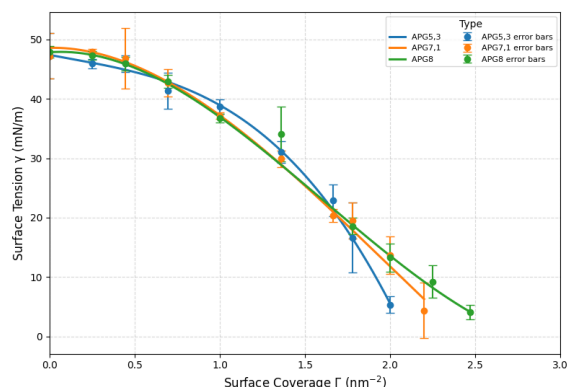


Figure 4. Interfacial tension against surface coverage plot for double-branched systems

For double-branched APG structures, both APG5,3 and APG7,1 reduce γ in a similar functional form, with clearer differences seen once Γ goes beyond 1 nm^{-2} . APG5,3 initially performs comparably to APG8, however past 1.66 nm^{-2} its curve falls more steeply. This can be attributed to the enhanced packing efficiency at high surface coverages, which arises from the wedge-like effective shape created by its symmetric branching. APG7,1, which has an asymmetric branch structure, produced slightly higher γ across most of the range, indicating less optimal packing and increased steric hindrance at the interface.

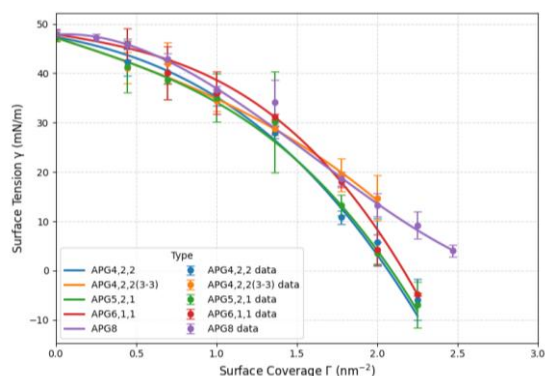


Figure 5. Interfacial tension against surface coverage plot for triple-branched systems

Triple-branched APGs exhibited the most distinct morphological trends relative to APG8. Although these structures are not commercially available, they serve as useful model systems for

exploring how extreme branching patterns influence interfacial behaviour. Among them, APG4,2,2—the most symmetric of the set—showed the steepest decrease in γ at high surface coverages, indicating that it organises efficiently once the interface becomes crowded. APG5,2,1 displays a similar trend, albeit slightly less pronounced, reflecting the marginally lower degree of symmetry in its branching pattern. In contrast, APG6,1,1 performs noticeably worse; its uneven, directional branching produces steric interference near the headgroup region, making it harder for the molecules to arrange effectively at the interface and resulting in consistently higher γ values.

APG4,2,2(3-3) departs from the behaviour of APG4,2,2 despite having the same number of branches. Because its branching occurs further down the tail, its interfacial organisation is altered, and the resulting γ profile reflects a different mode of adsorption compared to the headgroup-proximal isomer. A more detailed explanation for these distinctions is provided in the subsequent bulk-concentration analysis in section 5.3, where the effect of branching position on adsorption efficiency and CMC is examined directly.

Across all systems, the triple-branched variants outperform APG8 once the interface becomes crowded (around 1.778 nm^{-2}), showing lower γ values in the high-coverage regime. This suggests that branching can become advantageous under strongly packed interfacial conditions. Beyond approximately 2.0 nm^{-2} , several profiles exhibit irregular behaviour—including sharper drops, increased scatter, and occasional upward fluctuations—indicating that the interface is approaching supersaturation. These deviations signal the onset of the CMC region, where the interfacial layer can no longer accommodate additional surfactant in an orderly manner.

5.2 Identifying the CMC point

For each trajectory, the spatial density distribution $\rho(z)$ was computed across the simulation box and averaged over the equilibrated portion of the run. The resulting profiles at each surface coverage value display clear peak densities at the surface interfaces. These peak values are then plotted to reveal the maximum density values across the different surface coverage values. And following the methodology described in 4.2, the CMC point can be clearly

identified.

For all surfactants studied, the equilibrium density profiles exhibit a well-defined maximum at high surface coverage. This density peak corresponds to the onset of aggregation, signifying that the interface has become saturated and additional surfactants preferentially form bulk aggregates, signalling that the system has reached or surpassed the critical micelle concentration (CMC). With our MD simulations, the scale is too small to see micelles form and the interfaces past the CMC enter a metastable state where the surface interface continuously fluctuates, expands and ‘wrinkles’.

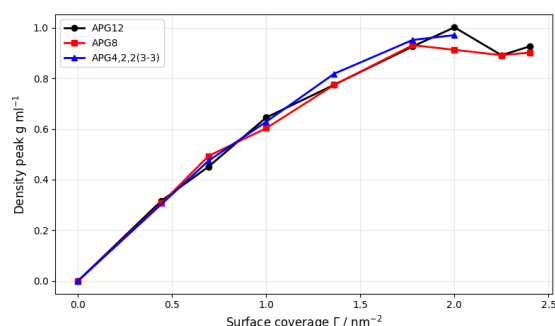


Figure 6. Density peaks against surface coverage for reference systems

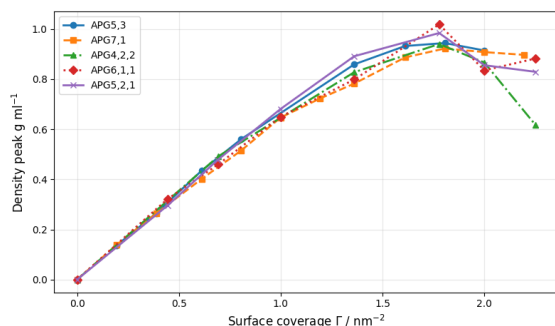


Figure 7. Density peaks against surface coverage for branched systems

Across all systems, the density–coverage profiles show that the maximum density peak reaches approximately $0.9\text{--}1.0\text{ g mL}^{-1}$ regardless of chain length or branching pattern. This indicates that the local packing structure at the interface is largely insensitive to molecular architecture: once adsorbed, all APGs organise into a similarly dense interfacial layer. The main difference lies not in the peak height but in the surface coverage required to reach this limit. APG12, with its longer hydrophobic tail, achieves saturation at $\Gamma \approx 2.0\text{ nm}^{-2}$, whereas APG8 and all of its branched derivatives converge at $\Gamma \approx 1.75\text{ nm}^{-2}$. This reflects the fact that longer tails promote stronger lateral hydrophobic attraction,

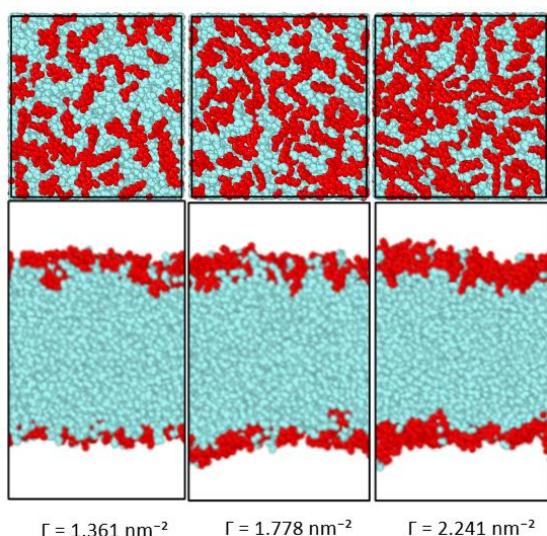
enabling tighter interfacial packing and therefore a higher Γ at saturation.

Although Γ cannot be used to define the CMC directly, these saturation points are useful because they indicate the maximum interfacial loading each surfactant can sustain before additional molecules must remain in the bulk or enter micellar aggregates. Thus, the density peak provides a physically meaningful marker for the onset of bulk aggregation, which corresponds to the CMC when translated into bulk concentration.

Among the branched derivatives, APG4,2,2(3–3) is the only structure that departs noticeably from the trend shared by the other APG8 isomers. While APG5,2,1, APG4,2,2, and APG6,1,1 all reach their maximum density at approximately the same surface coverage as APG8, the distal-branched APG4,2,2(3–3) reaches its density maximum at a higher Γ . This difference arises from its altered branching geometry. Relocating both branches to the far end of the tail reduces steric congestion near the headgroup—where packing constraints are most restrictive—and allows the molecule to orient and compress at the interface more efficiently than the other branched isomers. Because the molecule experiences less steric interference in the interfacial region, it can continue to accumulate at the interface for longer before packing limitations are reached. Consequently, the density peak occurs at a higher surface coverage, reflecting the fact that APG4,2,2(3–3) is able to accommodate more molecules at the interface before reaching its maximum packing state.

In this way, APG4,2,2(3–3) demonstrates that it is not only the *amount* of branching that matters but also its position: shifting branches toward the distal tail end enhances packing efficiency near the headgroup and enables the interface to sustain higher Γ before saturation.

The CMC can also be identified through a visual approach by inspecting snapshots at the interface using OVITO software. These snapshots in figure 8 capture the transition from a stable monolayer to an overcrowded, destabilised interface. At $\Gamma = 1.361 \text{ nm}^{-2}$, the interface remains flat and stable, with surfactants forming a well organised monolayer. At the CMC, $\Gamma = 1.778 \text{ nm}^{-2}$, slight interfacial wrinkling emerges, indicating the interface has reaching its maximum packing capacity. At $\Gamma = 2.241 \text{ nm}^{-2}$, the interface has become visibly distorted, with surfactant molecules starting to drift into the aqueous phase. **Figure 8.** OVITO visualisation of water-surfactant interface at three difference surface coverages.



5.3 Relating to Bulk Concentrations

Finally, when relating surface coverage data to bulk concentrations using the methodology outline in section 4.3, the following graphs were produced.

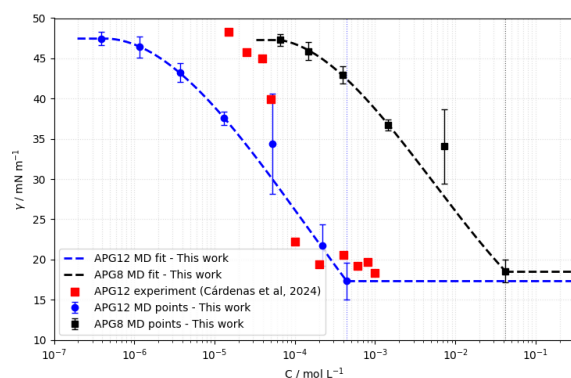


Figure 8. Interfacial tension of APG₁₂ and APG₈ as a function of bulk concentration with APG₁₂ experimental data [6]

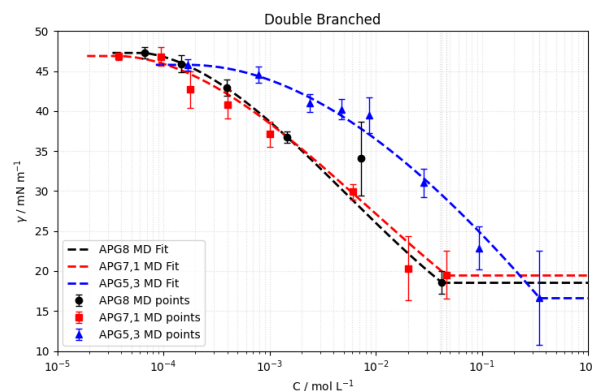


Figure 9. Interfacial tension of APG₈ and the double branched APG's as a function of bulk concentration.

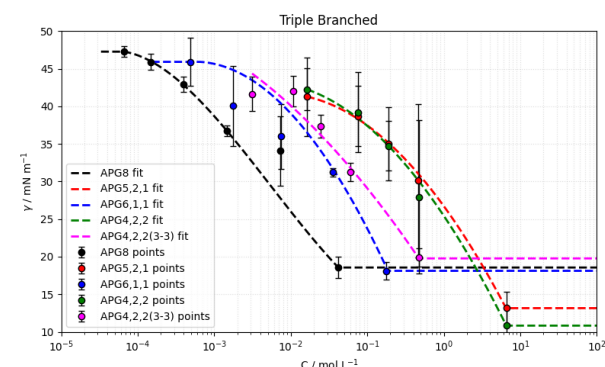


Figure 10. Interfacial tension of APG₈ and the triple branched APG's as a function of bulk concentration.

Across all systems, the interfacial tension decreases with increasing bulk concentration, consistent with classical non-ionic surfactant behaviour and Gibbs adsorption. At low concentration the interface is sparsely populated, but increasing bulk concentration promotes adsorption, leading to a progressive reduction in γ until a plateau is reached at the CMC. The good qualitative agreement between the MD curve for APG₁₂ and the experimental water/decane values of Cárdenas et al. [10] confirms that the simulations capture the essential adsorption physics, despite some overestimation of the initial γ .

The linear surfactant APG₈ provides the reference behaviour: its large sugar headgroup and slender unbranched C8 tail give it a predominantly cone-like molecular geometry, corresponding to a small packing parameter. This shape favours tight interfacial packing, producing low γ values and a relatively low CMC. APG_{7,1} exhibits a nearly identical interfacial tension profile and CMC, which can be rationalised by the fact that its single one-carbon branch does not significantly alter its cone-like geometry or effective packing parameter; hence the near-perfect overlap of the two curves.

In contrast, APG5,3 shows a more pronounced reduction in γ once adsorbed but a noticeably higher CMC. The addition of a second branch increases the tail's effective cross-sectional area, shifting the molecular shape toward a truncated cone with a larger packing parameter. This reduces adsorption efficiency, delaying interfacial saturation and causing the higher CMC, even though the molecule ultimately packs well enough at the interface to lower γ more than APG8 and APG7,1 at comparable concentrations.

All triple-branched surfactants display higher γ at low concentration and a slower descent toward the plateau compared to APG8. The increased steric bulk introduced by multiple branches pushes the packing parameter further toward the cylindrical regime, where dense interfacial packing becomes sterically hindered. As a result, all triple-branched species exhibit reduced adsorption efficiency and increased CMC values despite having the same C8 hydrophobic length as APG8.

Among them, APG6,1,1 shows the mildest effect of branching, with a γ_{CMC} similar to APG8 but occurring at a higher concentration, indicating delayed adsorption rather than diminished ultimate packing. APG5,2,1 and APG4,2,2 behave very similarly, both possessing bulky, highly branched tails that adopt near-cylindrical geometries. APG4,2,2 reaches a slightly lower γ_{CMC} , which is consistent with its larger hydrophobic volume from a longer second branch that enhances van der Waals contact once adsorbed.

The positional isomer APG4,2,2(3–3) clearly demonstrates how branching location can influence interfacial behaviour even when the number of branches is unchanged. Compared to APG4,2,2, shifting both branches from the headgroup-proximal position to the far end of the tail produces a distinct change in adsorption characteristics.

In the updated data, APG4,2,2(3–3) reaches its plateau at a lower bulk concentration, indicating a lower CMC relative to APG4,2,2. Relocating the branches distally reduces steric crowding near the headgroup—the region most relevant during the early stages of adsorption. This reduction in steric hindrance improves adsorption efficiency, enabling interfacial saturation to occur earlier.

Despite this shift in CMC, the surface

tension at the plateau remains very similar between APG4,2,2(3–3) and APG4,2,2. This indicates that both molecules ultimately pack to a comparable density once the interface is saturated. In other words, the branching position strongly affects the adsorption pathway, but it does not significantly alter the final interfacial packing state.

6 Conclusion and Outlook

The simulations conducted in this project highlight how minor alterations to the alkyl polyglucoside architecture can produce disproportionately large changes in macroscopic interfacial behaviour. More importantly, the collective trends revealed across all molecular families point toward broader design principles that extend beyond the individual systems studied. The results suggest that APG performance at water-oil interfaces is governed not simply by hydrophobicity or chain length, but by a more subtle interplay between molecular geometry, steric accessibility, and the way each architecture negotiates the interface during adsorption. This insight opens new conceptual routes for tailoring surfactants through structural manipulation rather than purely through compositional changes.

One of the most significant implications of this work is that surfactant efficiency can be modified predictably by adjusting branching pattern and branching position, offering a modular design strategy for sustainability-driven surfactant development. The fact that distal branching can partially recover interfacial performance that proximal branching disrupts suggests that synthetic efforts could decouple biodegradability from performance more effectively than current APG design strategies allow. This positions APGs as a versatile platform for iterative molecular engineering rather than a fixed family of commercially available molecules.

The study also highlights several avenues for future research. First, the relationship between the packing parameter and adsorption kinetics, although qualitatively inferred here, would benefit from a quantitative free-energy analysis to map out the energetic penalties associated with different branching motifs. Second, looking at the behaviour of these surfactants at curved interfaces, such as in emulsions, could prove essential for real-world applications. Third, integrating machine-learning models or coarse-grained simulations may allow rapid screening of hypothetical APG architectures, accelerating the identification of optimised

branching patterns without the cost of full atomistic simulations. An important next step would be to examine how the branching trends identified at the interface propagate into the micellar regime. Coarse-grained simulations offer a practical route to predict whether specific branching motifs favour spherical, wormlike, or lamellar aggregates, providing a deeper link between molecular architecture and mesoscale self-assembly. Establishing how branching influences micelle morphology would strengthen the connection between interfacial behaviour and bulk solution organisation, further informing the rational design of APG surfactants.

Finally, this project reveals the need for a more unified experimental–computational framework for APGs. While existing experimental datasets are sparse, the trends observed here demonstrate that simulations can supply mechanistic insight into why certain APGs behave as they do rather than merely reproducing the behaviour. With targeted experimental measurements, particularly around branching isomers that have not yet been synthesised, APG development could progress beyond empirical screening toward rational molecular design. Overall, this work does not simply characterise the interfacial properties of known APGs; it demonstrates that branching is a controllable molecular parameter with systematic, predictable consequences, and therefore a promising lever for engineering environmentally friendly surfactants with bespoke performance. The conclusions reached here provide a conceptual foundation for future APG innovation and motivate further exploration of structure–function relationships at fluid interfaces.

References

- [1] www.grandviewresearch.com. (n.d.). *Surfactants Market Size & Share Analysis Report, 2016-2022*. [online] Available at:
- [2] MarketsandMarkets (2022). *Surfactants Market Size & Forecast [Latest]*. [online] MarketsandMarkets. Available at:
- [3] Europa.eu. (2025). *Substance Information - ECHA*. [online]
- [4] Congruencemarketinsights.com. (2023). *Global Alkyl Polyglucosides (APG) Biosurfactants Market Size, Trends, Share, Growth, and Opportunity Forecast, 2023 - 2030 by Congruence Market*
- [5] Cárdenas, H., M. Ariif H. Kamrul-Bahrin, Seddon, D., Othman, J., Cabral, J.T., Andrés Mejía, Shahrudin, S., Matar, O.K. and Müller, E.A. (2024). Determining interfacial tension and critical micelle concentrations of surfactants from atomistic molecular simulations. *Journal of Colloid and Interface Science*, 674, pp.1071–1082. doi:
- [5] Bahrin, M.A.H.K. (2024). *Unveiling Structure-Property Relationships of Sugar-Based Surfactants at Fluid Interfaces: A Multiscale Approach Combining Atomistic and Coarse-Grained Molecular Simulations*. PDF. p.141
- [6] Z. Y. Liu, C. Wang, H. Zhou, Y. Wang, L. Zhang, L. Zhang, S. Zhao, *Characterizing the impact of surfactant structure on interfacial tension: a molecular dynamics study*, J. Mol. Model. 23 (2017) 112.
- [7] S. S. Jang, S. T. Lin, P. K. Maiti, M. Blanco, W. A. Goddard, P. Shuler, Y. Tang, *Molecular dynamics study of a surfactant-mediated decane–water interface: effect of molecular architecture of alkyl benzene sulfonate*, J. Phys. Chem. B 108 (2004) 12130–1214
- [8] B. Yoo, E. Marin-Rimoldi, R. Gotchy Mullen, A. Jusufi, E.J. Maginn, Discrete fractional component Monte Carlo simulation study of dilute nonionic surfactants at the air–water interface, Langmuir 33 (2017) 9793–9802
- [9] Thompson, A.P., Aktulga, H.M., Berger, R., Bolintineanu, D.S., Brown, W.M., Crozier, P.S., in 't Veld, P.J., Kohlmeyer, A., Moore, S.G., Nguyen, T.D., Shan, R., Stevens, M.J., Tranchida, J., Trott, C. and Plimpton, S.J. (2022). LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, 271, p.108171. doi:https://doi.org/10.1016/j.cpc.2021.108171.
- [10] Cárdenas, H., M. Ariif H. Kamrul-Bahrin, Seddon, D., Othman, J., Cabral, J.T., Andrés Mejía, Shahrudin, S., Matar, O.K. and Müller, E.A. (2024). Determining interfacial tension and critical micelle concentrations of surfactants from atomistic molecular simulations. *Journal of Colloid and Interface Science*, 674, pp.1071–1082. doi:

The Effects of Amino Acid Impurities and Well Conditions on Vancomycin Crystallisation

Katy Knoyle and Sisi Liu

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Crystallisation is a purification technique which can be used to purify vancomycin, an important pharmaceutical product used to treat severe bacterial infections. This paper focuses on finding more optimal conditions for the crystallisation of vancomycin and investigating the effects of two amino acids, lysine and histidine, on the formation of vancomycin crystals. Three buffer concentrations, two types of buffers, two types of salt and different ratios of amino acids are compared. Results show the optimal conditions tested for the crystallisation of vancomycin are 0.4 – 0.6 µg/µl of vancomycin, in 0.05 M of acetate buffer with NaCl as the salting-out agent. Lysine and histidine do not affect the shape of the octahedral vancomycin crystals, but they could still be interacting with the vancomycin molecules and could be part of the crystal structure.

1. Introduction

Vancomycin is a glycopeptide antibiotic used to treat gram-positive bacterial infections such as methicillin-resistant *Staphylococcus aureus* (MRSA) and methicillin-susceptible *Staphylococcus aureus* (MSSA) (Patel et al, 2024). MRSA is a type of bacteria that is resistant to most antibiotics. MSSA is a similar bacterium but less resistant to certain antibiotics (HSC Public Health Agency, n.d.). Both can live harmlessly on top of the skin but if they get under the skin through cuts or open wounds and left untreated, they can lead to many complications such as pneumonia, sepsis and can be life-threatening (NHS, 2023) (Centers for Disease Control and Prevention, 2024).

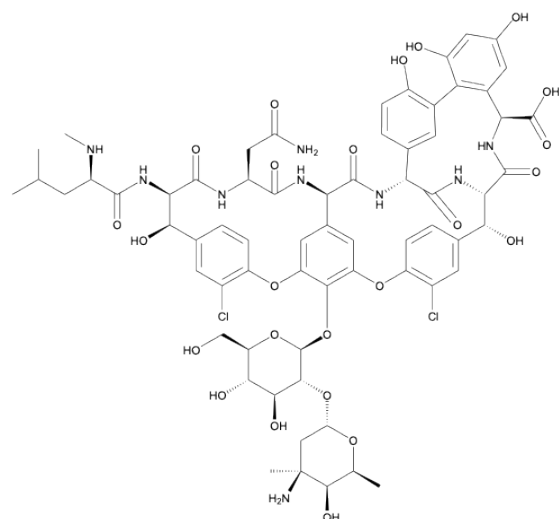


Figure 1 The molecular structure of vancomycin. Figure produced on ChemDraw

Chromatography is commonly used to purify pharmaceutical products in industry. However, chromatography uses large amounts of solvents

which increases costs and have a negative environmental impact. Purification costs can account for about 80% of production costs (Rathore et al., 2015). However, crystallisation can be used to purify vancomycin. Crystallisation is an important separation technique used in the production of pharmaceutical products. It is used for the purification of final products as well as for isolating and purifying starting materials and intermediates (Cotel et. al., 2020). Crystallisation is an effective way of obtaining high-purity pharmaceutical products in a cost-effective manner (Klyushnichenko, 2003).

The purpose of this research is to find the effects of well conditions, such as buffer choice, salt choice and vancomycin hydrochloride concentration, on the formation and growth of vancomycin crystallisation. This is to be able to help optimise the production of vancomycin crystals to lower costs and environmental impacts.

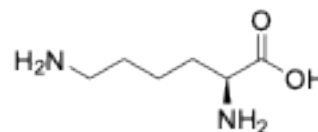


Figure 2 The molecular structure of lysine. Figure produced on ChemDraw

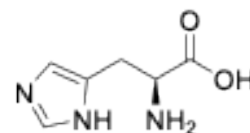


Figure 3 The molecular structure of histidine. Figure produced on ChemDraw

The second objective of this research is to determine if amino acids affect vancomycin crystallisation. Vancomycin has a molecular formula of C₆₆H₇₅Cl₂N₉O₂₄ (PubChem, 2020) and a

structure presented in Figure 1. Lysine has a molecular formula of $C_6H_{14}N_2O_2$ and histidine has a formula of $C_6H_9N_3O_2$ (Figure 2 and 3). Since all three of these molecules have amino groups (NH_2) that could be protonated to NH_3^+ and carboxyl groups ($COOH$) that can be reduced to COO^- , these molecules can interact with each other. Lysine and histidine are our chosen amino acids since they both have positively charged R-groups. Lysine is a strongly basic amino acid, because of the extra NH_2 group at the end, and histidine has an imidazole ring and is therefore a weakly basic amino acid.

2. Background

A few studies on vancomycin crystallisation have been completed prior to our work. Research has been undertaken to see how the crystallisation changes with different concentrations of vancomycin using NaCl as the salting-out agent. The study also compares the crystallisation at pH 2.6, 3.6, 4.6, 5.6 of sodium acetate. Many crystals habits are formed under these conditions, such as octahedral, stand-alone needle, oblongate leaf-like needle and sea urchin-like crystals (Pu, 2022). The research carried out and presented in this research project report uses NaCl as the salting-out agent and sodium acetate buffer but compares it to Na_2SO_4 and Bis-Tris buffer.

No papers were found on the effects of lysine and histidine on vancomycin crystallisation, which leaves scope for innovation for our research.

3. Methods and Materials

3.1 Materials

Sodium acetate (anhydrous, $\geq 99\%$), Bis-Tris ($\geq 98.0\%$), glacial acetic acid ($\geq 99.7\%$), methylene blue, L-Lysine ($\geq 98.0\%$), sodium chloride (NaCl, $\geq 99.0\%$), and sodium sulphate (Na_2SO_4 , $\geq 99.0\%$) were purchased from Sigma Aldrich. L-Histidine ($\geq 98\%$) was purchased from Thermo Scientific and hydrochloric acid (HCl, 37%) was purchased from VWR Chemicals. Vancomycin hydrochloride was obtained from Apollo Scientific. Deionised water was obtained from the laboratory. The chemicals were not further purified.

3.2 Initial Screening

Experiments were conducted to compare the formation of vancomycin crystals using two salting-out agents, sodium chloride (NaCl) and sodium

sulphate (Na_2SO_4). Vancomycin hydrochloride concentrations of 40 – 240 $\mu g/\mu l$ with 40 $\mu g/\mu l$ increments were used to be able to compare the crystals with different vancomycin concentrations. Two stock solutions of buffers were prepared: one 0.05 M acetate buffer of pH 5.39 and one 1 M Bis-Tris buffer of pH 9.91. These were prepared following the method by AAT Bioquest (2025). A Balance Technology balance, with an accuracy of $\pm 0.1 mg$ and Mettler Toledo FiveEasy pH meter, with an accuracy of ± 0.01 , were used to measure the required powder weight and to measure the pH, respectively, for all experiments conducted. A 1 M solution of NaCl and a 1 M solution of Na_2SO_4 were prepared for each buffer type. 0.1 g/mL vancomycin solutions in acetate and Bis-Tris buffer were also prepared. The salts and vancomycin were dissolved in each buffer using a magnetic stirrer.

A 48-well Swissci sitting drop plate was used for the initial screening. 4mL of buffer solution was pipetted into each well. Acetate was in the first four rows and Bis-Tris was pipetted into the final four rows. 40 – 240 $\mu g/\mu l$, in increments of 40 $\mu g/\mu l$ of vancomycin hydrochloride in buffer solution were pipetted into each well. 3-4 droplets of the salt with buffer solution (S+B) were pipetted into each reservoir. Then, 0 – 200 $\mu g/\mu l$, starting from 200 $\mu g/\mu l$ and decreasing in increments of 40 $\mu g/\mu l$ of salt were pipetted into each well. The wells with no salt were used as control wells. The amount of liquid in each well was kept at a constant of 4 μl . Since each salt were tested with each buffer and vancomycin hydrochloride concentration, and all wells had a repeat, this led to 48 different tests.

To prevent sample spillage and constant conditions, the wells were sealed with clear duct tape, from the brand Duck Tape and stored in an incubator at $25 \pm 1^\circ C$.

3.3 Methylene blue dye test

After eight days of incubation and observing what had grown in the wells of the initial screening, the duct tape was removed, and a few microns of methylene blue dye were pipetted into the wells. This was to determine if the crystals grown are vancomycin or salt crystals.

3.4 Condition Optimising

An experiment was conducted to compare the effects of buffer concentration on the vancomycin

crystals. Acetate and Bis-Tris buffer were used at the concentrations of 0.05 M, 0.5 M and 1 M. Stock solutions of 1 M of each buffer were prepared and then divided into three vials each so that the solutions could be diluted using de-ionised water to the required concentrations. The pH of each solution was measured using a Mettler Toledo FiveEasy pH meter. To keep the pH between the two buffers as similar as possible to each other, whilst keeping each buffer inside its ideal buffer range, hydrochloric acid and glacial acetic acid were used to lower the pH of the Bis-Tris and acetate buffers, respectively. The buffer range of Bis-Tris is pH 5.8 to 7.2 (Merck, n.d) and its pH 3.7 – 5.6 for acetate buffer (Dawson et al, 1959). Therefore, a final pH of 5.8 and 5.6 were measured for the Bis-Tris and acetate buffers, respectively.

Two 24-well Hampton Research VDX hanging drop plates with sealant were used for this experiment. Each reservoir was filled with 700 μm of NaCl in buffer solution. Instead of wells, like in sitting drop, solutions were pipetted on 22x22mm squares of 0.13-1.17mm thick cover glass. The two plates had identical conditions apart from one plate had acetate buffer and the other had Bis-Tris buffer. The salt concentration in each well was 0.4M and each well had a total of 4 μl of liquid.

3.5 Amino Acid Impurities

An experiment was conducted to investigate the effect of two amino acids, Lysine and Histidine on vancomycin hydrochloride crystallisation. For this sitting drop experiment, 1.6 μmol of NaCl was pipetted into each well along with 4 μl of 1 M Bis-Tris buffer. The vancomycin hydrochloride concentration is varied from 160 $\mu\text{g}/\mu\text{l}$ to 240 $\mu\text{g}/\mu\text{l}$ with 40 $\mu\text{g}/\mu\text{l}$ increments. For the two amino acids, the amino acid to vancomycin molar ratio tested was 2:1, 1:1, and a control experiment of 0:1 where no amino acids was added. To obtain neutrally charge amino acids, a buffer pH of 9.7 was chosen for lysine and pH 7.6 for histidine. Bis-Tris buffer was used for this experiment, rather than acetate buffer,

because pH 9.7 and 7.6 are within its optimal buffer range. The pH was adjusted by adding HCl.

3.6 Image Analysis

Each sample was analysed under an Olympus CX41 microscope using GXCAM HiChrome-MET and photos were taken daily using GX Capture to track the crystal formation and growth. Particle size and particle size distribution are crucial parameters for the application of active pharmaceutical compounds. Image analysis using microscopy is an important tool for monitoring the crystallisation process. Unlike alternative methods, including Nuclear Magnetic Resonance (NMR), and X-Ray Diffraction (XRD), using image analysis does not break the culturing environment, and continued analysis over time can be achieved. Image analysis is more cost-effective and environmentally friendly than alternatives since no extra solvents or equipment are required.

Fiji was used as the image analysis software. A scale of 66 pixels/100 μm ratio in 5x images used, provided by GX Capture. A minimum particle size of 10 μm was set to eliminate any dust from being captured. 500 μm was used as the maximum particle size to be captured, the default value from the Macro-V93 (Vancleef et al., 2022), and since none of the crystals are more than 500 μm . The parameters were tuned until an accurate capture was achieved.

4. Results and Discussion

4.1 Salt Comparison

No crystals were observed under any of the tested conditions with Na_2SO_4 . However, crystals were observed in the wells with NaCl. Sulphate (SO_4^{2-}) is more kosmotropic than chloride (Cl^-) in the Hofmeister series and therefore has a better ability to precipitate the peptide. Therefore, a lower SO_4^{2-} concentration is required before precipitation starts forming (Gregory et al., 2022). This explains why for the wells containing Na_2SO_4 , for both buffers, precipitate form in the wells instead of crystals.

Table 1 Structures observed in the wells before dyeing with methylene blue. Pictures have been taken with a 5x magnification.

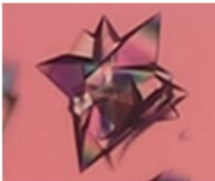
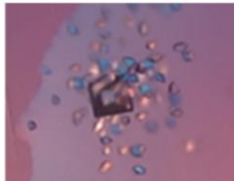
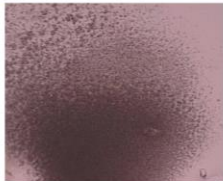
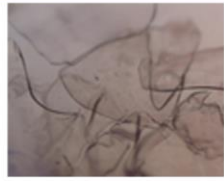
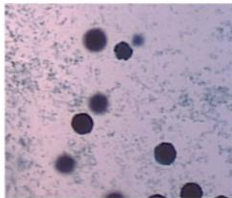
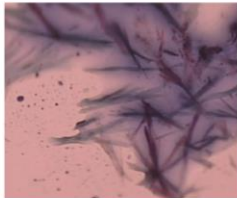
Octahedral (O)	Twinning Octahedral (TO)	Secondary Nucleation Octahedral (SO)	Precipitate (PPT)	'Ice' like
				
0.05M Acetate pH 5.6, NaCl 1.2µg V + 2.8µmol S + 4µl B (Day 7)	0.05M Acetate pH 5.6, NaCl 1.2µg V + 2.8µmol S + 4µl B (Day 4)	1M Bis-Tris pH 9.9, NaCl 1.2µg V + 2.8µmol S + 4µl B (Day 5)	0.05M Acetate pH 5.6, NaCl 0.4µg V + 3.6µmol S + 4µl B (Day 6)	0.05M Acetate pH 5.6, NaCl 1.2µg V + 2.8µmol S + 4µl B (Day 6)

Table 2 Structures observed in the wells before dyeing with methylene blue. Pictures have been taken using a 5x magnification.

Possible Methylene Blue Crystal	Standard Alone Needle (SAN)	Sea Urchin Needle (SUN)	Thick Needle (TN)	Urchin Thick Needle (UTN)
				
0.05M Acetate pH 5.6 NaCl (0.8µg V + 3.2µmol S + 4µl B) Day 14	0.05M Acetate pH 5.6 NaCl (1.2µg V + 2.8µmol S + 4µl B) Day 4	1M Bis-Tris pH 9.9 NaCl (1.6µg V + 2.4µmol S + 4µl B) 5 days	1M Bis-Tris pH 9.9 NaCl (0.8µg V + 3.2µmol S + 4µl B) 14 days	1M Bis-Tris pH 9.9 NaCl (1.2µg V + 2.8µmol S + 4µl B) 14 days

4.2 Methylene blue dye test

A methylene blue dye test was used to identify whether these crystals are peptide, and therefore vancomycin, or salt crystals. The dye stained the crystals blue and therefore it's suggested that these are vancomycin crystals.

After dyeing the crystals on day 8, different phenomenon were observed, as shown in Figure xxx. Many variables were changed whilst adding the dye and therefore there are a lot of uncertainties about the formations. In order to add the dye, the duct tape that was sealing the wells and reservoirs were removed and the solutions were exposed to the atmosphere. This meant that the samples were

exposed to dust or particles in the air and fluctuations in temperature, pressure and humidity which may have had effects on the crystal growth.

As seen in Table 2, after dyeing the crystals, individual needle crystals, clusters of needles that look like sea-urchins and thick needles were were formed, distinct from the octahedral-shaped crystals observed before. These needles could be vancomycin polymorphs. However, in all experiments tested, these were only observed in the wells that had been dyed and therefore are very likely to be caused by the change in conditions when the dye was added and could be acetate or Bis-Tris crystals, depending on which buffer the vancomycin was in, or dye crystals.

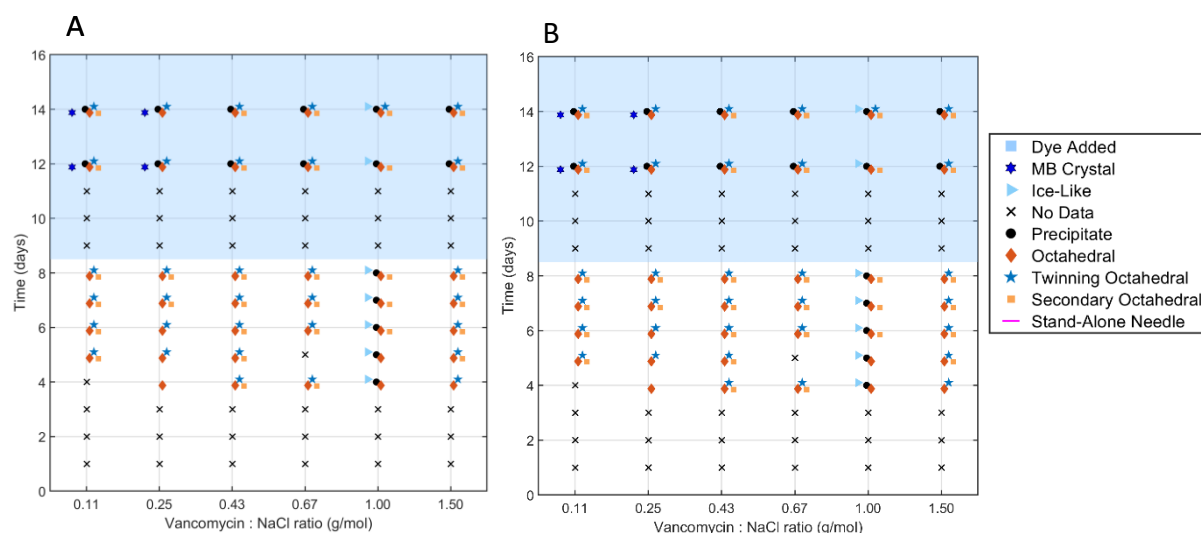
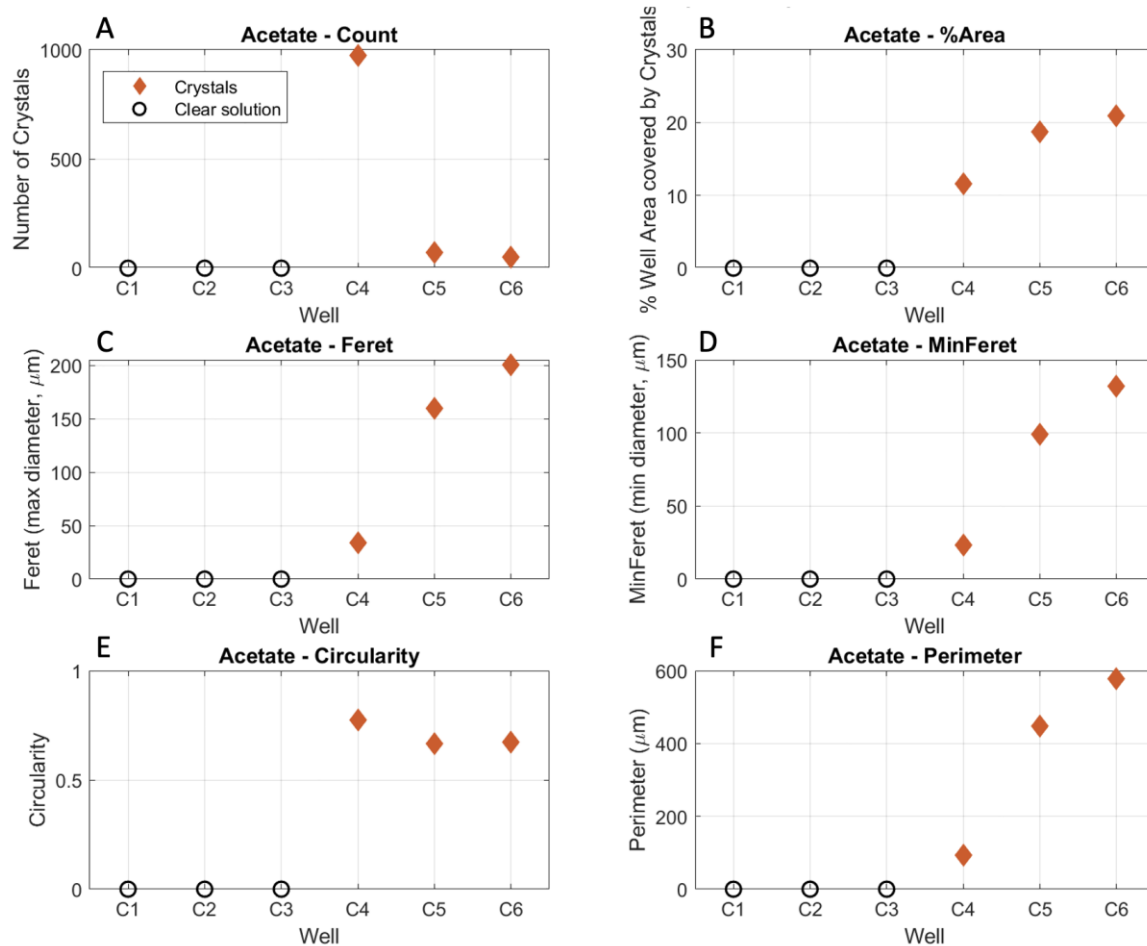


Figure 4 Structures observed for each day for each vancomycin to NaCl ratio. A is the first experiment and B is the repeat experiment.

4.5 Image Analysis



In some occasional cases, Fiji can't detect all crystals, for example crystal at upper level are captured but the ones below are not, it is usually due to low constraction, for this we go to process and enhance contrast by ~20% before the plugin.

Another case would require more effort is that, when the crystal are too crowded, and only crystals in certain part in the picture are clearly distinguishable, we crop this section out and use it as a representative for the whole well, and apply the normal steps for the cropped photo, in this case,

outcome of Feret, MinFere, perimeter and circularity are still reliable, but number of crystal need to be recalculated by work out the area percentage of cropped photo and the original one, then use the Count outcome of cropped photo divide this percentage. And the %surface covered is not as reliable.

4.3 Buffer Comparison

Using NaCl for the salting-out agent, crystals were formed in both acetate and Bis-Tris buffers. Although the trends between the wells using acetate and bistris as buffers remained similar, and both buffers can produce vancomycin crystals, there are a few differences between them.

However, different phenomenon were observed in the two buffers, as shown in Figure xxx and xxx. For both buffers, Octahedral crystals (O), sea-urchin like crystals (SUO), methyl blue crystals (MB), ice-like structures, precipitate (PPT), secondary nucleation crystals (SO) and standard individual alone needles (SAN) were observed.

4.3 Concentration Comparison

Figure xxxx shows what was observed in each well. The graph illustrates the formations with the salt-to-vancomycin ratio used (S:V) on the x-axis and the time in days on the y axis.

As results, we found that two buffer has similar performance, NaCl better performance, salt conc with 0.4-0.6M has best performance.

4.4 Amino Acid Impurities

Adding lysine and histidine did not prevent the crystals from forming. Octahedral crystals were observed in all the wells tested. Figure xxx shows the different classifications of crystals formed in the plates with the lysine and histidine impurities. There are three main classifications that our wells are divided into. A) A well full of closely spaced crystals, which completely cover the well. There is no room for further growth. These were formed in the wells with a 2:1 lysine : vancomycin ratio, for all three concentrations tried. From the octahedral shape, these crystals are likely to be pure vancomycin crystals. B) The middle of the drops in the wells affected all had a higher concentration of crystals in the centre but less the distance from the centre increased. B was observed in the control wells of vancomycin only, with no amino acids. It was also observed with the 1:1 lysine : vancomycin ratio. This implies that a small amount of lysine does not affect how many crystals grow. C) The crystals observed are more spaced out than A), meaning there is room for them to grow larger if..... This classification is observed in the two wells that contain histidine.

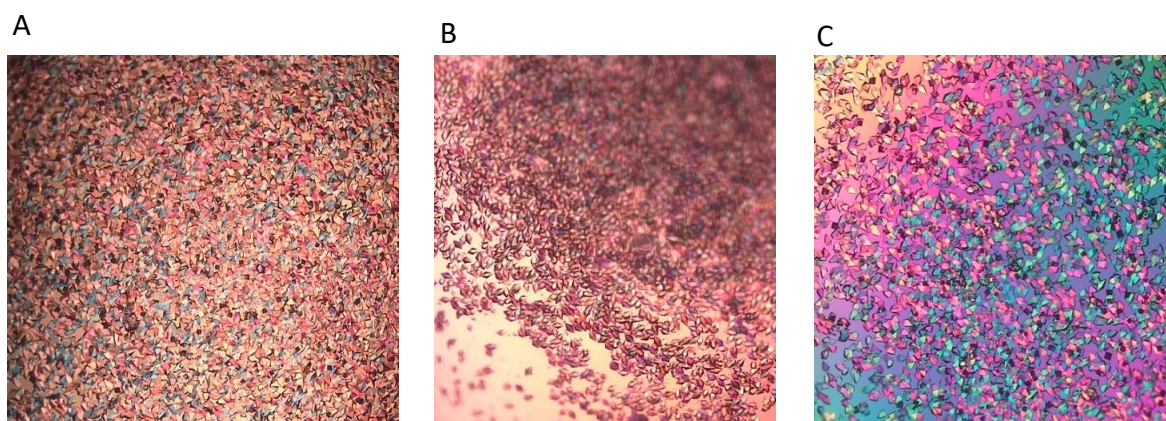


Figure 5 The classifications of the crystals in the wells

4.5 Particle Size Distribution

The crystals grown ranged between xx and xx μm . These are similar to the values of xxxx given in xxxx. All crystals grown are too small to be seen with the naked eye.

In theory, histidine and lysine could interact with the vancomycin molecule. Vancomycin has multiple sites hydrogen bonding can occur, such as the phenolic hydroxyl groups, amide oxygens and carboxylates, as highlighted in Figure xxx. Lysine has an amino (NH_3^+) group which can donate a proton and histidine has an imidazole ring which can be partially protonated and can be a hydrogen bond

donor or acceptor. This means that the lysine and histidine molecules can interact and form hydrogen bonds with the vancomycin molecules. Lysine's flexible aliphatic chain and histidine's compact imidazole ring can fill voids in the crystal lattice and increase the amount of order in the structure, making it more energetically stable.

Three image analysis tools were used to analyse the size distributions of the particles, CellProfiler, ImageJ and Fiji. CellProfiler is a biological image quantification, especially for cell based assays. It has high-throughput ability, but is not designed for crystal detection, so it can't sure what certain morphology should go for, and rounded captions happens especially If the boundary's not clear, partial area lost during detection. Then a Macro file was run with function: Feret, Minferet, self defined equivalent circular

Diameter. ImageJ is Good at identifying crystals and it is Lightweight and fast, it has Huge plugin ecosystem, finally Macro script (ImageJ Macro, Python, Java) enables automation. But wrong capture happened nearly in all crystals so to avoid the background being identified as crystal, the dark regions being missed, and the "crystal edges" being hollowed out, I added an HSB dual-threshold union, ROI restriction, and closing/hole-filling operations. It solves the problem of hollow crystal but at the same time, the crystal boundaries became more blurred, and the reflections caused by the polariser being misclassified as background also remain unresolved.

However, due to the research aim should focus on impurities analysis, this self-defined macro is not fully refined. Therefore, an existing open-source Fiji method is applied.

To further analyse the crystals formed in the solutions with the amino acids, they should be sent to a synchrotron. Although the crystals formed look the same as the vancomycin crystals formed without the amino acids, there are a lot of channels in the crystals where the impurities could be present. An analysis with the synchrotron would show whether these amino acids are present inside the crystal structure.

Fiji was used for image analysis of the plates. For the three wells of data in the acetate buffer, trends are observed. For the low concentrations of vancomycin, no octahedral crystals were formed,

and therefore there is no data. As the vancomycin concentration increases, the ferret diameter

5 Conclusions and Outlook

From the initial screening, it can be concluded that Bis-Tris and acetate both work as buffers to facilitate vancomycin crystal growth. However, the experiments conducted imply that acetate is more predictable and therefore is more likely to be able to be controlled to get a desired output. NaCl is a better salting-out agent than Na₂SO₄, since no crystals are formed using Na₂SO₄. From the conditions tried, the optimal vancomycin concentrations are 160-240 µg/µl since these concentrations produce vancomycin octahedral crystals. In the optimisation plate, no crystals are produced in the wells with a vancomycin concentration lower than this. There is potential for lysine and histidine to interact with the vancomycin and be present in the crystal. However, further research should be conducted to determine whether lysine and histidine are present in the crystals obtained.

There is scope for additional research in the field of vancomycin crystallisation. Research should be conducted on the optimal conditions to grow bigger crystals, and crystals One potential improvement to be made would be to use a silica nucleation site slide for hanging drop experiments. Induction time would be interesting to observe to see the effects of the variables discussed in this report.

In the cropped image analysis scenario, when the crystal are too crowded, and only crystals in certain part in the picture are clearly distinguishable, we crop this section out and use it as a representative for the whole well, and apply the normal steps for the cropped photo, in this case, outcome of Feret, MinFeret, perimeter and circularity are still reliable, but number of crystal need to be recalculated by work out the area percentage of cropped photo and the original one, then use the Count outcome of cropped photo divide this percentage. A potential improvement would be apply a enhancement factor on the number obtained, because crowded part should have more crystal compare to the clearly distinguished part. And sometimes when the cropped picture is at the edge, a even higher enhancement factor should be applied here, since there's more background in the picture at edge of droplet. And the %surface covered is not reliable.

6 Acknowledgements

The authors would like to thank Hirokatsu Yoshioka and Priyam Chakraborty for their continued support and guidance throughout the project.

7 References

AAT Bioquest, Inc.

(2025). *Quest Calculate™ Acetate Buffer (pH 3.6 to 5.6) Preparation and Recipe*. AAT Bioquest. Available at: <https://www.aatbio.com/resources/buffer-preparations-and-recipes/acetate-buffer-ph-3-6-to-5-6>.

AAT Bioquest, Inc. (2025). *Quest Calculate™ Bis-Tris Buffer (1 M, 6.5 pH) Preparation and Recipe*. AAT Bioquest. Available at: <https://www.aatbio.com/resources/buffer-preparations-and-recipes/bis-tris-buffer-ph-6-5>. <https://www.aatbio.com/resources/buffer-preparations-and-recipes/bis-tris-buffer-ph-6-5>

Centers for Disease Control and Prevention (2024). *Methicillin-resistant Staphylococcus aureus (MRSA) Basics*. [online] Methicillin-resistant Staphylococcus aureus (MRSA). Available at: <https://www.cdc.gov/mrsa/about/index.html>

Cote, A., Erdemir, D., Girard, K.P., Green, D.A., Lovette, M.A., Sirota, E. and Nere, N.K. (2020). Perspectives on the Current State, Challenges, and Opportunities in Pharmaceutical Crystallization Process Development. *Crystal Growth & Design*, 20(12), pp.7568–7581
doi:<https://doi.org/10.1021/acs.cgd.0c00847>
https://pubs.acs.org/doi/pdf/10.1021/acs.cgd.0c00847?ref=article_openPDF

Dawson, R. M. C., Elliott, D. C., Elliott, W. H. and Jones, K. M. (1959) Data for Biochemical Research. Available at: https://ia600207.us.archive.org/14/items/in.ernet.dli.2015.142415/2015.142415.142415.Data-For-Biochemical-Research_text.pdf

Gregory, K.P., Elliott, G.R., Robertson, H., Kumar, A., Wanless, E.J., Webber, G.B., Craig, V.S.J., Andersson, G.G. and Page, A.J. (2022). Understanding specific ion effects and the Hofmeister series. *Physical Chemistry Chemical Physics*, 24(21), pp.12682–12718.
doi:<https://doi.org/10.1039/d2cp00847e>.

HSC Public Health Agency (n.d.). *Meticillin-Sensitive Staphylococcus aureus (MSSA)*. Available at:

<https://www.publichealth.hscni.net/sites/default/files/2019-05/MSSA%20information%20factsheet.pdf>

Klyushnichenko, V. (2003). Protein crystallization: from HTS to kilogram-scale. *Current opinion in drug discovery & development*, [online] 6(6), pp.848–54. Available at: <https://pubmed.ncbi.nlm.nih.gov/14758756/>

Merk (n.d.). *Buffer Reference Centre*, [online].

Available at:

<https://www.sigmaaldrich.com/GB/en/technical-documents/protocol/protein-biology/protein-concentration-and-buffer-exchange/buffer-reference-center>

NHS (2023). *MRSA*. [online] NHS. Available at: <https://www.nhs.uk/conditions/mrsa/>

Patel, S., Preuss, C.V. and Bernice, F.

(2024). *Vancomycin*. [online] National Library of Medicine. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK459263/>

Pu, S. and Hadinoto, K. (2022). Salting-Out crystallization of glycopeptide Vancomycin: Phase behavior study to control the crystal habit. *Chemical Engineering Science*, [online] 262, p.118057.
doi:<https://doi.org/10.1016/j.ces.2022.118057>

PubChem (2020). *Vancomycin*. [online] PubChem. Available at: <https://pubchem.ncbi.nlm.nih.gov/compound/Vancomycin>

Rathore, A., Pathak, M., Ma, G. and Bracewell, D. (2015). Re-use of Protein A Resin: Fouling and Economics. *UCL Discovery (University College London)*.

Vancleef, A., Maes, D., Van Gerven, T., Thomassen, L.C.J. and Braeken, L. (2022). Flow-through microscopy and image analysis for crystallization processes. *Chemical Engineering Science*, 248, p.117067.
doi:<https://doi.org/10.1016/j.ces.2021.117067>

Orthogonal Collocation on Finite Elements Method for Simulation of Membrane Contactors in CO₂ Capture

Jack Burke

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

This paper implements a predictive mathematical model of a high-pressure membrane contactor in Python. The work uses an existing steady-state model developed by Quek et al.^[1], consisting of partial differential mass balance equations, along with phase coupling algebraic equations. The governing equations are discretised using an orthogonal collocation on finite elements method, yielding a system of nonlinear algebraic equations that can be computed by a Newton solver. The model is then validated against previous work including a sensitivity analysis for CO₂ absorption with respect to key design and operational variables and showing feasibility for an OCFE model for MBC simulation.

Key Words: Natural gas sweetening, CO₂ absorption, Membrane based contactor, Spectral Method, Orthogonal Collocation, Finite Element Method

1. Introduction

Natural gas (NG) is widely used for power generation, heating and a broad range of industrial processes. NG is currently the third most used fossil fuel for energy consumption, and the third overall^[2]. Despite the share of renewables steadily increasing since the 2000s, the demand of global energy is still too high to be satisfied without the use of fossil fuels. Consequently, NG usage is projected to rise 32% by 2050^[3].

Raw NG often does not meet sales-gas specifications, so it must therefore be “sweetened” before it can be used, a process that usually removes CO₂ and H₂S. For commercial natural gas, the allowable CO₂ limit has been reported to be 2.5%^[4]. Liquefied natural gas (LNG) is a lot more stringent, typically having an upper limit of 100 ppm^[5]. There are several operational and technical reasons to do this; CO₂ and water in the natural gas react to form carbonic acid, causing corrosion of the pipeline. CO₂ reduces the global heating value of NG^[6], making it less efficient during combustion. In LNG, the effects are more severe, as higher quantities of CO₂ will freeze and cause damage to the cryogenic equipment.

The current industry standard is a tall packed absorption column, in which the NG is contacted counter currently with an amine solvent over a large gas-liquid interfacial area, thereby enhancing CO₂ mass transfer from the gas phase into the liquid. However, to achieve the contact time to reach NG specifications, very tall columns^[7] are required, and are subject to issues such as flooding and foaming.

Membrane-based contactors (MBCs) apply the same principle without phase dispersion; a porous membrane keeps the gas phase and liquid solvent separate, and the membrane acts as the mass transfer interface. MBCs often have a much greater specific surface area, offering 1000 m²/m³^[8], whereas packed columns typically only offer

an active interfacial area of 100-250 m²/m³^[8], however pressure drop constraints and membrane porosity often reduce the effective interfacial area for MBCs.

MBCs lack simple empirical correlations to aid with design and have transport resistances depending on the solvent, membrane properties, compressibility, wetting and reaction phenomena. A mathematical model is therefore required to understand the relationships between design parameters and MBC performance. The governing equations for an MBC consist of coupled, nonlinear partial differential equations (PDEs) with additional algebraic equations (AEs) enforcing phase equilibrium and coupling. Given the steep concentration gradients occurring around phase edges, this produces a stiff system, where the solution may not be stable for certain numerical methods.

Classic, straightforward numerical techniques such as finite differences require lots of elements in order to have a high degree of accuracy, and lead to large systems of equations with high computational cost, and can only output a discrete solution. In addition, without a very small step size, steep gradients are poorly captured in these solutions.

Orthogonal Collocation on Finite Elements (OCFE) offers an effective numerical method for this problem type. In OCFE, the spatial domain is split up into finite elements, and within each element the solution is approximated by a polynomial. The governing PDEs are then enforced at a set of collocation points chosen according to orthogonal polynomials, allowing them to be expressed as AEs suitable for standard Newton solvers. This choice sets the accuracy: for a fixed number of collocation points, the polynomial degree controls the truncation error, which usually drops very quickly as degree increases, needing minimum number of points is required for high accuracy.

The study pursues the following objectives. First, to implement a mathematical model of a hollow-fiber MBC

in Python using the discretisation method of OCFE. The gas mixture to be simulated will be a CO₂-N₂ mix and the amine solvent will consist of a MDEA-PZ mix (methyldiethanolamine and piperazine) The OCFE model is then scrutinised by determining if the modification of variable input parameters produces consistent with those observed experimentally, showcasing the robustness of the

model. These parameters will include solvent flow rate, gas flow rate, inlet gas pressure, loading effects and other operational conditions acting on the modules removal efficiency and absorption flux. Then, a brief discussion benchmarking against other discretisation schemes, demonstrating how OCFE can deliver improved accuracy for reduced computational costs in MBC simulation.

2. Background

Nomenclature	
Symbols A_m – membrane area C – concentration D – diffusivity E – removal efficiency F – volumetric flowrate H – Henry's constant L – Lagrangian basis M – molar flowrate P – pressure R – gas constant T – temperature Z – compressibility factor $f_{CO_2,l}^{in}$ – CO₂ loading n_f – number of fibers r₁ – radius of tube r_w – wetted radius r₂ – radius at which shell starts r₃ – radius at which shell ends (outer radius of model) r_m – inner radius of module	v – velocity z – axial length position δ – pore radius δ_w – critical pore radius δ_{a,b} – Kronecker delta ε – porosity γ – surface tension μ – dynamic viscosity ρ – density τ – tortuosity θ – contact angle ω – wetting ratio φ – PSD distribution function η – packing density Subscripts in – at inlet out – at outlet g – in gas phase l – in liquid phase sol – solvent

This section summarises the steady-state hollow-fiber MBC model developed by Quek et al.^[1], excluding the energy balance.

The governing equations for the concentration and gas velocity are described by the following mass balance, where the overall domain of a fiber has been split into the following partitions: tube ($0 \leq r \leq r_1$), dry-membrane ($r_1 \leq r \leq r_w$), wet-membrane ($r_w \leq r \leq r_2$), liquid ($r_2 \leq r \leq r_3$). Here, r₃ is related to r₂ by:

$$r_3 = \frac{r_2}{\sqrt{\eta}} \quad (0)$$

where η is the packing density. The constants in these equations were taken from the reported values of Quek et al.

The governing equations describe the coupled IVP of CO₂ concentration and gas velocity in the tube, radial diffusion and reaction in the dry-membrane and wet-membrane regions, and convection–diffusion–reaction in the liquid phase, together with the corresponding boundary conditions at the gas–membrane, gas–liquid, and liquid–

membrane interfaces.

Membrane properties and pressure-drop correlations are also included to capture the variation of local per-slice pressure and wetting behavior. The wetted radius r_w represents the average depth of liquid penetration into the membrane pores and is defined via the wetting ratio, as introduced in Eqs. (22)–(23).

Gas Phase, (r, z) ∈ [0, r₁] × [0, L]:

$$\frac{d\overline{C_{CO_2}}}{dz}(z) = -2 \frac{\varepsilon D_{CO_2,s}}{\tau r_1 \overline{v}_g(z)} \left[1 - \frac{ZRT_g}{P_g} \right] \frac{\partial C_{CO_2}}{\partial r}(r_1^+, z) \quad (1)$$

$$\frac{d\overline{v}_g}{dz}(z) = -2 \frac{\varepsilon D_{CO_2,s}}{\tau r_1} \frac{ZRT_g}{P_g} \frac{\partial C_{CO_2}}{\partial r}(r_1^+, z) \quad (2)$$

Gas Inlet, z = L:

$$\overline{v}_g(L) = \frac{M_g^{in}}{n_f \pi r_1^2 \rho_g}, \quad \overline{C_{CO_2}}(L) = \frac{y_{CO_2,g}^{in} P_g}{ZRT_g} \quad (3, 4)$$

Gas-Membrane Interface, $r = r_1$:

$$\bar{C}_{CO_2}(z) = C_{CO_2}(r_1^+, z) \quad (5)$$

Membrane-Dry Phase, $(r, z) \in [r_1, r_w(z)] \times [0, L]$:

$$0 = \frac{\partial^2 C_{CO_2}}{\partial r^2}(r, z) + \frac{1}{r} \frac{\partial C_{CO_2}}{\partial r}(r, z) \quad (6)$$

Gas-Liquid Interface, $r = r_w(z)$:

$$ZRT_g C_{CO_2}(r_w^-(z), z) = H_{CO_2,l} C_{CO_2}(r_w^+(z), z) \quad (7)$$

$$D_{CO_2,g} \frac{\partial C_{CO_2}}{\partial r}(r_w^-(z), z) = D_{CO_2,l} \frac{\partial C_{CO_2}}{\partial r}(r_w^+(z), z) \quad (8)$$

$$0 = \frac{\partial C_{sol}}{\partial r}(r_w^+(z), z) \quad (9)$$

Membrane-Wet Phase, $(r, z) \in [r_w(z), r_2] \times [0, L]$:

$$0 = \frac{\varepsilon}{\tau} D_{i,l} \left[\frac{\partial^2 C_i}{\partial r^2}(r, z) + \frac{1}{r} \frac{\partial C_i}{\partial r}(r, z) \right] + R_i(C_i(r, z)) \quad (10)$$

Liquid-Membrane Interface, $r = r_2, i \in \{CO_2, sol\}$:

$$C_i(r_2^-, z) = C_i(r_2^+, z), \frac{\varepsilon}{\tau} \frac{\partial C_i}{\partial r}(r_2^-, z) = \frac{\partial C_i}{\partial r}(r_2^+, z) \quad (11,12)$$

Liquid Phase, $(r, z) \in [r_1, r_w(z)] \times [0, L], i \in \{CO_2, sol\}$:

$$v_l(r) \frac{\partial C_i}{\partial z}(r, z) = D_{i,l} \left[\frac{\partial^2 C_i}{\partial r^2}(r, z) + \frac{1}{r} \frac{\partial C_i}{\partial r}(r, z) + \frac{\partial^2 C_i}{\partial z^2}(r, z) \right] + R_i(C_i(r, z)) \quad (13)$$

Liquid Boundary, $r = r_3, i \in \{CO_2, sol\}$

$$\frac{\partial C_i}{\partial r}(r_3, z) = 0 \quad (14)$$

Liquid Inlet, $z = 0, r \in [r_2, r_3]$:

$$C_{sol}(r, 0) = C_{sol}^{in}, \quad C_{CO_2}(r, 0) = f_{CO_2,l}^{in} C_{sol}^{in} \quad (15,16)$$

Liquid Outlet, $z = L, r \in [r_2, r_3], i \in \{CO_2, sol\}$:

$$\frac{\partial C_i}{\partial z}(r, L) = 0 \quad (17)$$

$$v_l(r) = 2\bar{v}_l(r_3^2 - r_2^2) \frac{r^2 - r_2^2 + 2r_3^2 \ln \frac{r_2}{r}}{4r_2^2 r_3^2 - r_2^4 - 3r_3^4 - 4r_3^4 \ln \frac{r_2}{r}} \quad (18)$$

$$\bar{v}_l = \frac{F_l}{n_f \pi r_3^2 (1 - \eta)} \quad (19)$$

Membrane Properties and Pressure Drop, $z \in [0, L]$:

$$\frac{\Delta P(z)}{\Delta z} = P_l^{out} - P_g + 8\mu_l \bar{v}_1 \frac{r_2^2 - r_1^2}{4r_2^2 r_3^2 - r_2^4 - 3r_3^4 - 4r_3^4 \ln \left(\frac{r_2}{r_3} \right)} (L - z) \quad (20)$$

$$\int_{\delta_w(z)}^{\delta^{max}} \delta^2 \varphi(\delta) d\varphi = \omega(z) \int_0^{\delta^{max}} \delta^2 \varphi(\delta) d\delta \quad (22)$$

$$r_w = r_2 - [r_2 - r_1] \omega(z) \quad (23)$$

$$\Delta P_{TMPD}(z) = P_l^{out} - \frac{P_l}{\Delta z} (L - z) + V \rho_l g (L - z) - P_g^{in} \quad (24)$$

$$\delta_w(z) = \frac{2\gamma \cos \theta}{\Delta P_{TMPD}(z)} \quad (25)$$

3. Methods

3.1 Membrane and real gas characteristics

Quek et al.^[1] modelled the pore size distribution as fitting a log pore-normal distribution ($\varphi(\delta)$), and a similar approach was used in this paper. To ease computational time of calculating the analytical wetting ratio in Equation (22), a scaled and shifted logistic function surrogate was regressed in the form:

$$\omega(\delta_w) = \frac{(1 + \tanh(k_0 + k_1 \left(\frac{\delta_w}{\delta}\right) + k_2 \left(\frac{\delta_w}{\delta}\right)^2 + k_3 \left(\frac{\delta_w}{\delta}\right)^3))}{2} \quad (21)$$

where the constants were found to be $k_0 = 7.980$, $k_1 = -13.5$, $k_2 = 7.737$, $k_3 = -1.800$.

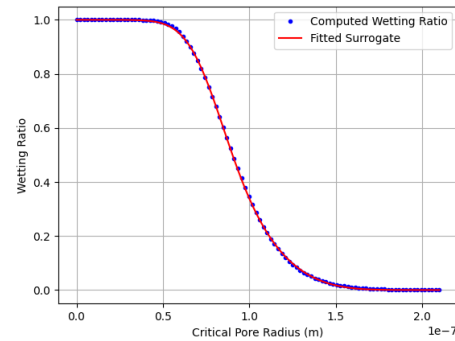


Figure 1: Surrogate fit of the wetting ratio against the critical pore radius

To account for the real-gas effects of the CO_2 at high pressure, a binary Peng-Robinson equation of state model was used to compute the compressibility factor, Z . Nitrogen was chosen as the second binary compound for sample simulations.

Given y_{CO_2} is decreasing down the length of the column, the expectation is the compressibility factor of the mixture will move towards that of pure N_2 ; the overall Z will become more ideal. This is verified in Figure 2.

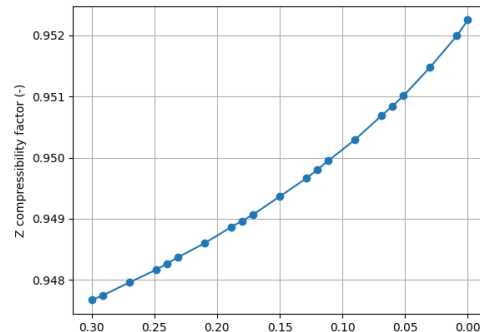


Figure 2: Variation of the compressibility factor Z throughout the axial length for a sample

3.2 Discretisation and Grid

The $[r_1, r_3] \times [0, L]$ domain is divided into three regions: the dry membrane, the wetted membrane, and the liquid region. A coordinate rescaling is applied only in the dry and wetted membrane subdomains so that the node grid is rectangular. All three regions are then subdivided further by the placement of equally spaced grid points N_{1w}, N_{w2}, N_{23} , and N_z respectively, denoting the numbers of grid points in the $[r_1, r_w]$, $[r_w, r_2]$, $[r_2, r_3]$, and $[0, L]$ intervals such that the grid points also cover the endpoints. For reference, it will be assumed that $N_{1w} = N_{w2} = N_{23} = N_z = N$. Note the gas phase is isomorphic to the first column of nodes in the dry membrane grid, avoiding the need for a separate gas phase mesh.

Elements are then defined by placing interior collocation points at the extrema of N_c order Chebyshev polynomials, given by $T'_{N_c-1}(x) = 0$, where $T_{N_c}(x)$ are the Chebyshev polynomials of the first kind, on $-1 \leq x \leq 1$, and by including the endpoints as extrema.

$$x_j = \cos\left(\frac{\pi j}{N_c - 1}\right), j = 0, \dots, N_c - 1 \quad (22)$$

These generate N_c Chebyshev-Gauss-Lobatto (CGL)^[9] nodes in each element. This effect of using a collection of low order orthogonal polynomials reduces computational complexity opposed to using one high order polynomial for an entire region. CGL nodes were specifically chosen to reduce Runge's phenomenon, specifically with the steep gradients incurred by the reaction term, and were determined easier to precompute than Legendre-Gauss-Lobatto points.

The indexing notation is defined as $C_{sp, (r_i, z_j)}^{[I, J]}$, where $[I, J]$ determine the overall subdomain index, and each point within the subdomain is referred to by index (r_i, z_j) , where $i, j \in 1, \dots, N_c \times 1, \dots, N_c$, $I, J \in \mathbb{N}$, refers to the radial and axial directions, respectively. With this notation, each element and subsequent collocation point within can be identified. The collocation points are then placed such that the indexing satisfies:

Dry & Wet Membrane Phase, $(r, z) \in [r_1, r_2] \times [0, L]$:	
$C_{CO_2, (z_j)}^{[0, J]}$	$J \in [1, (N-1)]$
$v_{g, (z_j)}^{[0, J]}$	$J \in [1, (N-1)]$
$C_{species, (r_i)}^{[I, J]}$	$(I, J) \in [1, 2(N-1)] \times [1, (N-1)(N_c - 2) + N]$

Table 1: Collocation points and indexing in membrane phases

where species $\in CO_2, \text{sol}$.

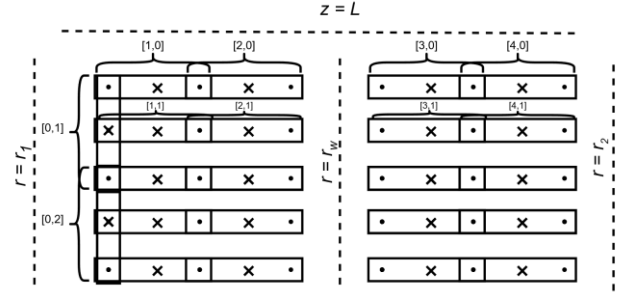


Figure 3: Schematic of elements in membrane region, $N = 2, N_c = 3$

Figure 3 shows a schematic drawing of elements in the wetted and dry membrane regions, illustrating the indexing convention. Note how the elements are 1-d in these partitions.

Liquid Region, $(r, z) \in [r_2, r_3] \times [0, L]$:	
$C_{species, (r_i, z_j)}^{[I, J]}$	$[I, J] \in [2N-1, 3N-2] \times [1, (N-1)]$

Table 2: Collocation points and indexing in the liquid phase $z = L$

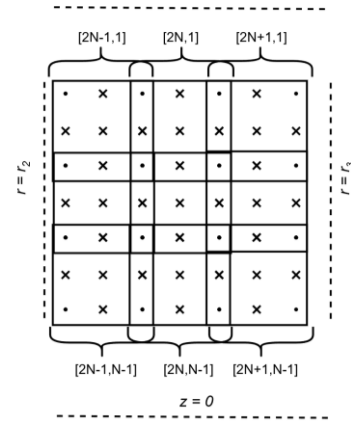


Figure 4: Schematic of elements in liquid region, $N = 3, N_c = 3$. N has not been substituted for clarity purposes.

Figure 4 shows a schematic drawing of elements in the liquid region, demonstrating the indexing. Note how the elements are 2-d in this region.

In the first element $[1, 0]$ in the membrane partition, a polynomial that passes through the CGL nodes, $\mathcal{C}^T = [C_{CO_2, (r_1)}^{[1, 0]}, \dots, C_{CO_2, (r_{N_c})}^{[1, 0]}]$ is interpolated as the Lagrange polynomial in the reference element $[-1, 1]$. For convenience, this is rewritten in matrix format:

$$\mathcal{C}(\mathbf{r}) = \mathbf{L}(\mathbf{r})\mathbf{C} \quad (23)$$

Where $\mathcal{C}(\mathbf{r})$ is the scalar interpolated value, $\mathbf{L}(\mathbf{r})$ is the row vector of Lagrange basis polynomials evaluated at $\mathbf{r}$, and $\mathbf{C}$ are the CGL nodal values:

$$\mathbf{L}(\mathbf{r}) = [L_1(\mathbf{r}) \quad \dots \quad L_{N_c}(\mathbf{r})] \quad (24)$$

$$L_i(r) = \prod_{m=1, m \neq i}^{N_c} \frac{r-r_m}{r_i-r_m} \quad (25)$$

For the CGL distribution of nodes, a differentiation matrix $\mathbf{D}$ of size N_c by N_c can be precomputed^[10] such that:

$$\mathbf{C}' = \mathbf{D}\mathbf{C} \quad (26)$$

where $\mathbf{C}'$ contains $\frac{\partial C}{\partial r}$ evaluated at CGL nodes.

$$c_i = \begin{cases} 2, & i = 1 \text{ or } i = N_c \\ 1, & \text{otherwise} \end{cases} \quad (27)$$

$$\mathbf{D}_{ij} = \begin{cases} (c_i/c_j)(-1)^{i+j}/(r_i - r_j), & i \neq j \\ -\sum_{m=1, m \neq i}^{N_c} \mathbf{D}_{im}, & i = j \end{cases} \quad (28)$$

The second derivative matrix $\mathbf{D}_2$ is then simply:

$$\mathbf{D}_2 = \mathbf{D}^2 = \mathbf{D}\mathbf{D} \quad (29)$$

For any physical element, $r \in [r_a, r_b]$ a mapping is applied from the reference coordinate $\xi \in [-1, 1]$ via $r(\xi) = \frac{r_b-r_a}{2}\xi + \frac{r_a+r_b}{2}$, and the differentiation matrices are scaled accordingly. This allows the reformulation of the governing PDE equations (6) and (10) as algebraic equations for the interior points. A similar approach in the z direction for the gas phase subdomain allows (1) and (2) to be also expressed as AEs.

In the first subdomain in the liquid region, $C_{sp,(r,z)}^{[2N-1,1]}$ as the governing PDE depends on both the r and z partials, the following is defined:

$$\mathbf{C} = \begin{bmatrix} C_{sp,(r_1,z_1)}^{[2N-1,1]} & \cdots & C_{sp,(r_{N_c},z_1)}^{[2N-1,1]} \\ \vdots & \ddots & \vdots \\ C_{sp,(r_1,z_{N_c})}^{[2N-1,1]} & \cdots & C_{sp,(r_{N_c},z_{N_c})}^{[2N-1,1]} \end{bmatrix} \quad (30)$$

$$\mathbf{L}_r(\mathbf{r}) = [L_1(r) \quad \cdots \quad L_{N_c}(r)] \quad (31)$$

$$\mathbf{L}_z(\mathbf{z}) = [L_1(z) \quad \cdots \quad L_{N_c}(z)] \quad (32)$$

From this, the scalar $C(\mathbf{r}, \mathbf{z})$ is the continuous tensor-product interpolant:

$$C(r, z) = \sum_{k=1}^{N_c} \sum_{l=1}^{N_c} \mathbf{C}_{kl} \mathbf{L}_k^z(z) \mathbf{L}_l^r(r) \quad (33)$$

$$C(r, z) = \mathbf{L}_z(z) \mathbf{C} \mathbf{L}_r(r)^T \quad (34)$$

For the nodal values, the differentiation matrices can be defined:

$$\left. \frac{\partial C}{\partial r} \right|_{r_l, z_k} = \sum_{m=1}^{N_c} \mathbf{L}_m^r(r_l) \mathbf{C}_{km} \delta_{z,k} = ((\mathbf{D}_r)(\mathbf{C}_{k,:}))^T_l \quad (35)$$

$$\left. \frac{\partial C}{\partial z} \right|_{r_l, z_k} = \sum_{m=1}^{N_c} \mathbf{L}_m^z(z_k) \mathbf{C}_{ml} \delta_{r,l} = ((\mathbf{D}_z)(\mathbf{C}_{:,l}))_k \quad (36)$$

where $\mathbf{C}_{:,l}$ and $\mathbf{C}_{k,:}$ refer to the k -th row and l -th column respectively and $\delta_{z,k}$ refers to the Kronecker delta for the k -th term in the $\mathbf{L}_z$ basis polynomials vector.

The second partial derivatives can therefore be computed in the same manner. Following the same remapping procedure as before, (13) can now be expressed as an AE.

3.1.1 Continuity Conditions

At the interfaces of elements within regions, continuity between the Lagrange interpolating polynomials is enforced. The concentration function is continuous by virtue of sharing a single point between subdomains.^[11]

Continuity of the derivative requires the flux of the previous element end to be equivalent to the next elements start. For the 1-d elements, the conditions are as follows:

Continuity Condition	Index
$\frac{\partial(c_{sp,(r_{N_c})}^{[I-1,J]})}{\partial r} = \frac{\partial(c_{sp,(r_1)}^{[I,J]})}{\partial r}$	$(I, J) \in [2, 2(N-1)] \times [1, (N-1)(N_c-2) + N]$
$\frac{\partial(c_{CO_2,(z_{N_c})}^{[0,J-1]})}{\partial z} = \frac{\partial(c_{CO_2,(z_1)}^{[0,J]})}{\partial z}$	$J \in [2, (N-1)]$
$\frac{\partial(v_{g,(z_{N_c})}^{[0,J-1]})}{\partial z} = \frac{\partial(v_{g,(z_1)}^{[0,J]})}{\partial z}$	$J \in [2, (N-1)]$

Table 3: Continuity conditions and indexing in the liquid phase

For the 2-d elements, the flux needs to agree between vertical and horizontal element interfaces. To determine the grid point continuity where 4 elements meet, a single condition combining both the horizontal and vertical flux is added. The conditions are as follows:

Continuity Condition	Index
$\frac{\partial(c_{sp,(r_{N_c},z_j)}^{[I-1,J]})}{\partial r} = \frac{\partial(c_{sp,(r_1,z_j)}^{[I,J]})}{\partial r}$	$(I, J) \in [2N, 3N-2] \times [1, (N-1)]$ $i \in [1, N_c - 1]$
$\frac{\partial(c_{sp,(r_i,z_{N_c})}^{[I,J-1]})}{\partial z} = \frac{\partial(c_{sp,(r_i,z_1)}^{[I,J]})}{\partial z}$	$(I, J) \in [2N-1, 3N-2] \times [2, (N-1)]$ $i \in [1, N_c - 1]$

$\left(\frac{\partial \left(c_{sp, (r_{N_c}, z_1)}^{[I,J]} \right)}{\partial z} - \frac{\partial \left(c_{sp, (r_{N_c}, z_{N_c})}^{[I,J-1]} \right)}{\partial z} \right) = \left(\frac{\partial \left(c_{sp, (r_1, N_c)}^{[I,J]} \right)}{\partial r} - \frac{\partial \left(c_{sp, (r_{N_c}, z_{N_c})}^{[I-1,J]} \right)}{\partial r} \right)$	$(I, J) \in 2N, 3N-2] \times [2, (N-1)]$
---	--

Table 4: Continuity conditions and indexing in the liquid phase

Elements at the boundary involving a very steep initial gradient can have the effect of distorting the derivative at the end of the element due to its global effect causing an oscillatory effect at the edge. In such a case, a more local method of calculating the derivative is employed, using the immediate points at the edge of the element.

3.1.2 Boundary Conditions

The remaining set of points to identify define the boundaries of the subregions. For this model, the entire region enclosed by r_1 and r_3 is formulated as a BVP. However, for the gas phase equations constitute an IVP there is no boundary condition at $z = 0$. The last point is therefore treated as an interior collocation point to approximate the gas solution, and is therefore acted on by the governing PDE operator. The resulting boundary conditions for all phases and interfaces are summarised in Table 5.

Boundary Condition	Index	Eq. No
$C_{CO_2, (z_1)}^{[0,1]} = \frac{y_{CO_2, g}^{in} P_g}{ZRT_g}$	N/A	(4)
$v_{g, (z_1)}^{[0,1]} = \frac{M_g^{in}}{n_f \pi r_1^2 \rho_g}$	N/A	(3)
$C_{CO_2, (r_1)}^{[1,j]} = C_{CO_2, (z_j)}^{[0,j]}$	$(j, J) \in [1, (N-1)(N_c-2) + N] \times [1, (N-1)]$	(5)
$C_{CO_2, (r_{N_c})}^{[I-1,J]} = \frac{H_{CO_2, l}}{ZRT_g} C_{CO_2, (r_1)}^{[I,J]}$	$(I, J) \in [N] \times [1, (N-1)(N_c-2) + N]$	(7)
$D_{CO_2, g} \frac{\partial C_{CO_2}^{[I-1,J]}}{\partial r}(r_{N_c}) = D_{CO_2, l} \frac{\partial C_{CO_2}^{[I,J]}}{\partial r}(r_1)$	$(I, J) \in [N] \times [1, (N-1)(N_c-2) + N]$	(8)
$0 = \frac{\partial C_{sol}^{[I,J]}}{\partial r}(r_1)$	$(I, J) \in [N] \times [1, (N-1)(N_c-2) + N]$	(9)

$C_{sol}^{[I,J]}(r_i, z_j) = 0$	$(I, J, i, j) \in [1, 2(N-1)] \times [1, (N-1)(N_c-2) + N] \times [1, N_c] \times [1, N_c]$	N/A
$\frac{\partial C_{sp}^{[I-1, (J-1)N_c+j]}}{\partial r}(r_{N_c}) = \frac{\partial C_{sp}^{[I,J]}}{\partial r}(r_1, z_j)$	$(I, J, j) \in [2N-1, 3N-2] \times [1, N-1] \times [1, N_c]$	(11)
$C_{sp}^{[I-1, (J-1)N_c+j]}(r_{N_c}) = C_{sp}^{[I,J]}(r_1, z_j)$	$(I, J, j) \in [2N-1, 3N-2] \times [1, N-1] \times [1, N_c]$	(12)
$\frac{\partial C_{sp}^{[I,J]}}{\partial r}(r_{N_c}, z_j) = 0$	$(I, J, j) \in [3N-2] \times [1, N-1] \times [1, N_c]$	(14)
$\frac{\partial C_{sp}^{[I-1,J]}}{\partial z}(r_i, z_1) = 0$	$(I, J, i) \in [2N, 3N-1] \times [1] \times [1, N_c]$	(17)
$C_{sp}^{[I,J]}(r_i, z_{N_c}) = C_{sp}^{in}$	$(I, J, i) \in [2N, 3N-1] \times [N-1] \times [1, N_c]$	(15,16)

Table 5: Boundary conditions in all phases

3.1.3 Solving

With these equations, the model can be built piecewise from its elements, where the layout for each individual element would be assembled piecewise:

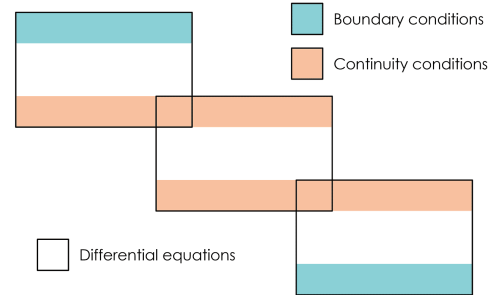


Figure 5: Diagram demonstrating assembly of the discretisation to the solver, for a 1-d set of elements^[12]

In Figure 5, the sparsity of the system is observed, as most of the PDE collocation points only depend on equations within the local element, as opposed to a global collocation method where each point would depend on every collocation point in the domain. The model was implemented in CasADi^[13] using its Python interface, and the resulting algebraic system was converted into residual expressions and solved at the collocation points with a Newton solver. The set of rewritten PDEs and conditions in Tables 3,4,5 were implemented in the form of:

$$F(\mathbf{x}_k) = \mathbf{r}_k \quad (37)$$

where $\mathbf{r}_k$ is the residual vector, $\mathbf{x}_k$ is a possible solution to the system, and function F is the system of equations. The Newton method therefore takes the form:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - J(\mathbf{x}_k)^{-1}F(\mathbf{x}_k) \quad (38)$$

where $\|\mathbf{r}_k\|$ (the Euclidean norm) is forced within a small tolerance, and J is the Jacobian of F .

An initial guess ($\mathbf{x}_0$) of a uniform concentration of order of magnitude $\sim C_{\text{mdea}}^{\text{in}}$ was used to ensure the solution would converge to be positive and accurate. Convergence was typically achieved given the system was using physically feasible parameters. The output from the solver was then interpolated using Lagrange polynomials and rescaled to reflect the wetting effects on the membrane.

The Jacobian calculated by CasaDi's AD is a sparse square matrix of dimension: $(9(N-1)(N_c-2) + N)^2 + ((N-1)(N_c-2) + N)$ per side. However, due to the fact OCFE only couples local variables inside each element, the majority of entries will be 0 for any given residual equation. As CasaDi stores sparse matrices in compressed sparse column format (CSC), operations involving the Jacobian only involve the nonzero values, drastically reducing memory usage. This allows for the execution time to be proportional to the number of nonzero variables, and not the number of Jacobian entries (See Figure x).

4. Results and Discussion

This multiphase problem is formulated in dimensional variables, velocity and concentration, and other operational

and physical constants such as pressure and diffusivity. In this study, robustness alongside computational cost will be the main factors discussed from the results of this model. The simulation was run on an AMD Ryzen™ 9 9800X3D (octa-core) CPU with 32GB of RAM, however, the hardware was underutilised, capping out around 15-20% utilisation likely due to the ineffectiveness of parallelism with CasADi's rootfinder.

For trends involving comparing model parameters, two metrics are defined, first the absorption flux [$\text{mol m}^{-2} \text{s}^{-1}$]:

$$\Phi = \frac{\pi R_m^2 \eta}{A_m} (C_{\text{CO}_2}(L)v_g(L) - C_{\text{CO}_2}(0)v_g(0)) \quad (39)$$

Then the removal efficiency, E [-]:

$$E = 1 - \frac{C_{\text{CO}_2}(0)v_g(0)}{C_{\text{CO}_2}(L)v_g(L)} \quad (40)$$

4.1 Convergence and Error

The absolute tolerance set for the solver was set to 10^{-4} . OCFE relies on coupled polynomials between the collocation points, and Douglas and Dupont^[14] showed that the error was proportional to $\mathcal{O}(N^{-5})$ when quartic functions were used in each element together with using the orthogonal polynomial collocation points. This is opposed to $\mathcal{O}(N^{-2})$ when using finite difference points within each element.

4.2 Reliability and Robustness

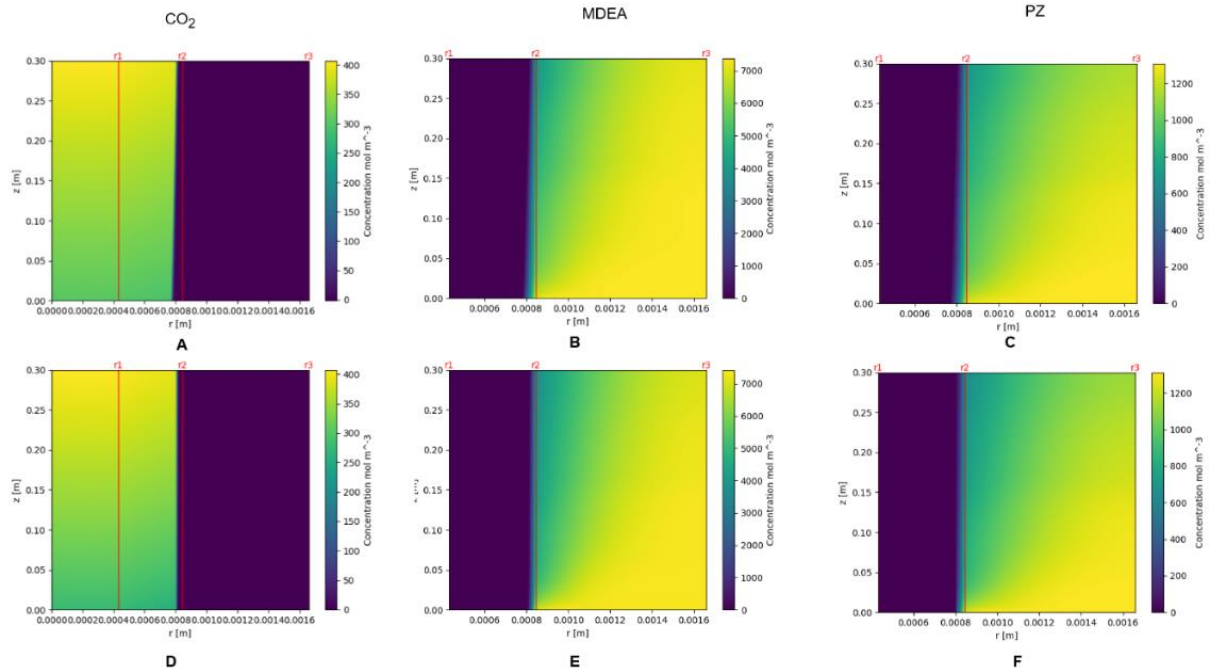


Figure 6: Sample concentration profiles of CO_2 , MDEA and PZ respectively, for vertical operation (A, B, C) and horizontal operation (D, E, F)

The sample concentration profiles showcase various features of the model. The gas flow is shown to be fully plug flow, whereas the liquid flow has variation in the shell due to the radial and axial diffusion effects shown by the 2-d equations. There is a steep concentration gradient in the wetted membrane, where only radial diffusion and reaction occur. It is noted that PZ reacts away faster than MDEA due to a higher macroscopic reaction term and diffusivity value.

From Figure 6, the difference in the vertical and horizontal operation of the MBC can be identified. In horizontal operation, $V = 0$ meaning the liquid static head will no longer contribute to the transmembrane pressure difference, leading to a larger δ_w value. As a result, much fewer pores will be fully wetted, i.e have a $\delta > \delta_w$, therefore the wetting radius appears to be much smaller and constant down z . This change is more pronounced when the module orientation is varied in Figure 7, where 0 represents a horizontal module, and $\frac{\pi}{2}$ rad represents a vertical module. As the orientation approaches $\frac{\pi}{2}$, the wetting distribution spreads out and the pores become more wetted and the CO_2 removal efficiency decreases as expected, as the CO_2 diffusion resistance will increase.

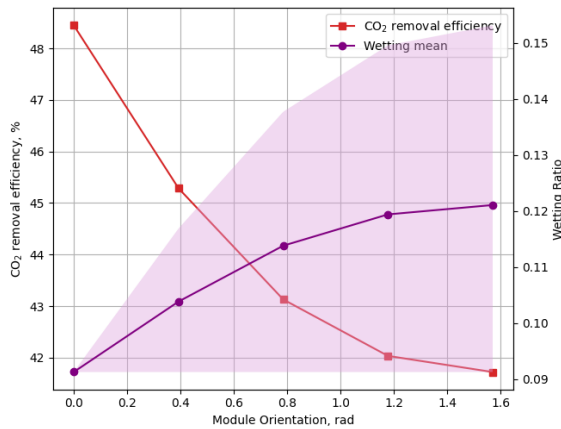


Figure 7: Module Orientation against CO_2 removal efficiency and spread of wetting values

Next, the effect of CO_2 loading is investigated. It is expected for the removal efficiency to decrease, as increased CO_2 at the inlet reacts with and consumes more solvent, meaning there is less free amine to react with the CO_2 downstream in the module. This is verified in Figure 8.

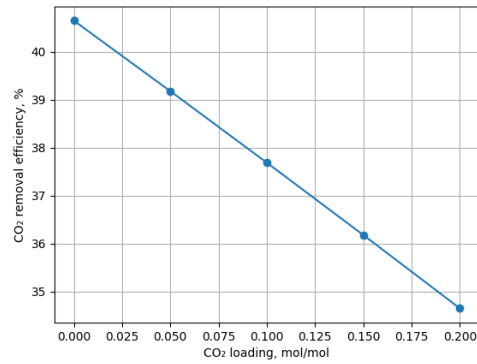


Figure 8: CO_2 loading in amine solvent against CO_2 removal efficiency.

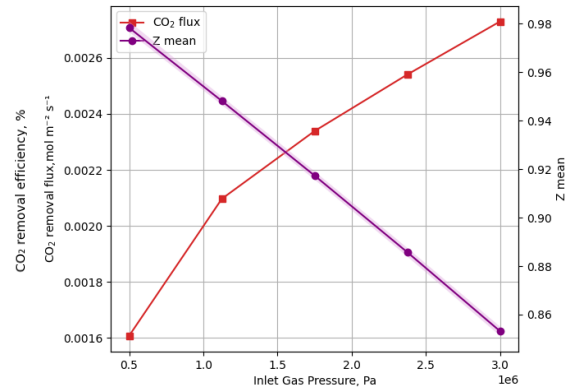


Figure 9: Inlet gas pressure against the compressibility factor Z and the CO_2 removal flux.

In Figure 9 it is observed the compressibility factor will decrease from ideality as the inlet gas mixture becomes less ideal with higher pressure. The increased pressure increases the gas phase CO_2 concentration and therefore increases the interfacial gradient $\frac{\partial c_{\text{CO}_2}}{\partial r}(r_1^+, z)$; since CO_2 is expected to react almost instantaneously away in the wetted membrane and thus the average flux increases due to a greater gas driving force. It is noted the wetting effects will remain the same due to the transmembrane pressure being held constant.

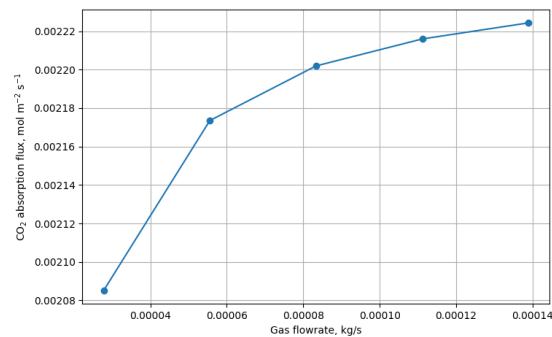


Figure 10: Gas mass flowrate against CO_2 absorption flux

The trends in Figure 10 and Figure 11 are to be expected given the previous results; however the impact is much more marginal on the removal flux. An increased gas mass flowrate increases the driving force similarly as seen to the increase in inlet gas pressure; however the overall removal efficiency can decrease if a smaller relative amount of the overall CO_2 is removed.

The increase in liquid solvent flowrate allows for more reaction to occur due to larger amounts of free amine being available to react.

The contact angle effect is shown in Figure 12. It is defined as the angle a liquid makes where it meets the solid surface, determined in the breakthrough pressure equation, Eq. (41).

$$P_c = \frac{-2\gamma\cos\theta}{\delta} \quad (41)$$

A larger contact angle causes the membrane to act more hydrophobically and requires a higher pressure for the liquid to penetrate and wet the pores. When the pores are less wetted (more gas filled) the gas mass-transfer resistance is decreased, allowing a higher absorption flux. As wetting decreases, the CO_2 absorption flux increases until wetting is completely suppressed.

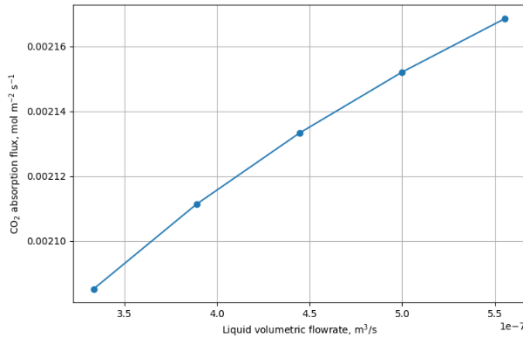


Figure 11: Liquid volumetric flowrate against CO_2 removal flux

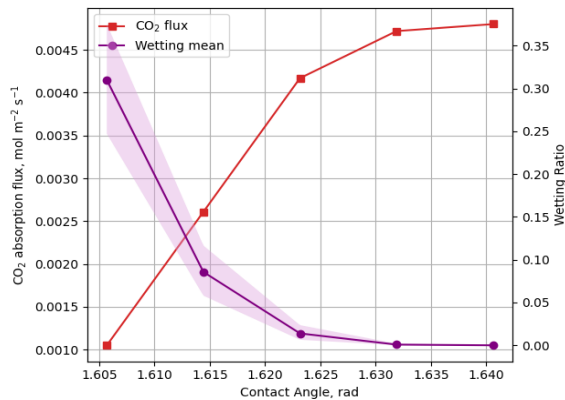


Figure 12: Contact angle against wetting values spread and CO_2 removal flux

4.3 Computational Cost

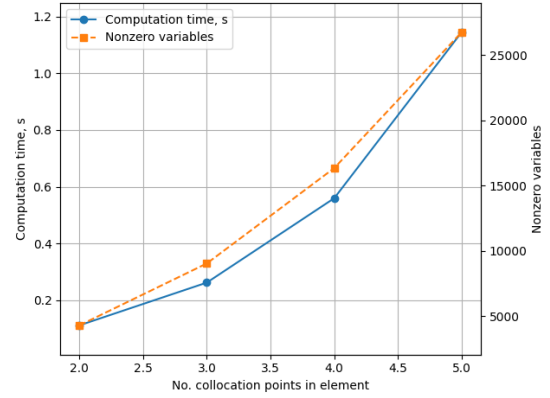


Figure 13: Number of collocation points inside an element against the computational time and NNZ

Most simulations were performed with the total collocation points set to $N_c = 5$, with 4 nodes per element. This basis was chosen as a compromise between modest runtime and accuracy, noting that higher values of N_c disproportionately increase the runtime for relatively small gains in accuracy. The dominant computational cost of the simulation is the nonlinear Newton solve, in which a large sparse Jacobian system is solved at each iteration. In the present implementation, this linear system is factorised using CasADi's CSpase linear solver, which performs a sparse LU factorisation of the Jacobian. Since each degree of freedom couples only to a fixed stencil of neighbours (see Figure ()), the number of nonzero entries in the Jacobian, grows linearly with the number of unknowns, and the cost of the CSpase factorisation and solves scales approximately with the number of nonzero values in the Jacobian. Assembling the system is comparatively negligible, as the spectral differentiation matrices are precomputed once and reused throughout the simulation.

As discussed in 4.1, for the same accuracy, a much less fine mesh is needed as is for a similar low order finite-difference scheme, i.e for a given accuracy orthogonal collocation will be a much faster method. A similar second-order finite difference scheme implemented by Quek et al.^[1] in gPROMS took on the magnitude of several minutes to complete, showing a huge improvement in computational time. Moreover, the spectral discretisation provides a continuous solution representation, in contrast to the pointwise nature of low-order finite-difference methods.

5. Conclusions

This work presented a steady-state hollow-fiber membrane contactor model for high-pressure CO_2 removal from natural gas, based on the formulation of Quek et al.^[1] and discretised using OCFE with Chebyshev–Gauss–Lobatto nodes. The model captures key physical features, including two-dimensional transport in the liquid shell, steep

reaction–diffusion gradients in the wetted membrane, and reproduces expected trends in CO₂ absorption flux and removal efficiency under variations in pressure, flow rates, loading, contact angle and module orientation. The resulting OFCE framework provides a practical and efficient tool for analysing the coupled effects of wetting, mass transfer, and operating conditions in MBCs quickly and accurately. Future extensions include embedding this OFCE model within a gradient-based optimisation analysis, where the design and operation variables are treated as the decision variables. As the discretised model can be used to produce

semi-continuous performance curves, sensitivities of the removal flux with respect to these variables can be calculated using CasADi automatic differentiation, allowing for optimal parameter design. In addition, a comparison of the model to experimental works to further verify the model should be completed. Finally, experimenting with different collocation points (Legendre) and solvers (Newton-Krylov) to investigate the impact of memory and execution time could be scope for further optimisation.

5. Bibliography

- [1] Quek, V.C., Shah, N. and Chachuat, B. (2018). Modeling for design and operation of high-pressure membrane contactors in natural gas sweetening. *Chemical Engineering Research and Design*, 132, pp.1005–1019. doi:<https://doi.org/10.1016/j.cherd.2018.01.033>.
- [2] Ritchie, H. and Rosado, P. (2020). *Energy Mix*. [online] Our World in Data. Available at: <https://ourworldindata.org/energy-mix>.
- [3] Gecf.org. (2025). *Global Gas Outlook 2050*. [online] Available at: <https://www.gecf.org/Publications-Data/Global-Gas-Outlook-2050> [Accessed 5 Dec. 2025].
- [4] Mazzetti, M.J., Skagestad, R., Mathisen, A. and Eldrup, N.H. (2014). CO₂ from Natural Gas Sweetening to Kick-start EOR in the North Sea. *Energy Procedia*, 63, pp.7280–7289. doi:<https://doi.org/10.1016/j.egypro.2014.11.764>.
- [5] www.fluxys.com. (n.d.). *LNG Specification Requirements*. [online] Available at: <https://www.fluxys.com/en/natural-gas-and-biomethane/supplying-europe/belgium/lng-belgium/lng-specification-requirements>.
- [6] Wasiu, A.B., Aziz, Abd.R.Abd. and Heikal, M.R. (2012). The Effect of Carbon Dioxide Content-natural Gas on the Performance Characteristics of Engines: A Review. *Journal of Applied Sciences*, [online] 12(23), pp.2346–2350. doi:<https://doi.org/10.3923/jas.2012.2346.2350>.
- [7] Udara S P R Arachchige, Mohsin, M. and Melaaen, M.C. (2012). Optimized Carbon Dioxide Removal Model for Gas Fired Power Plant. *European Journal of Scientific Research*, [online] 86, pp.348–359. Available at: https://www.researchgate.net/publication/331564299_Optimized_Carbon_Dioxide_Removal_Model_for_Gas_Fired_Power_Plant?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Ii9kaXJlY3QiLCJwYWdlIjojX2RpcmVjdCJ9fQ.
- [8] Hoff, K.A. and Svendsen, H.F. (2014). Membrane contactors for CO₂ absorption – Application, modeling and mass transfer effects. *Chemical Engineering Science*, 116, pp.331–341. doi:<https://doi.org/10.1016/j.ces.2014.05.001>. https://www.researchgate.net/publication/256529296_Transient_Chemical_Reaction_Analysis_by_Orthogonal_Collocation
- [9] Combette, A. (2023). *The Chebyshev interpolation* | Andrea Combette. [online] Andrea Combette. Available at: <https://andrea-combette.com/post/spectral-chebyshev/> [Accessed 5 Dec. 2025].
- [10] Itoh, S. (2019). *Derivation of Chebyshev Differentiation Matrix*. [online] Available at: https://micronanopi.net/Chebyshev/CDM/Derivation_CDM_B.pdf [Accessed 5 Dec. 2025].
- [12] Weitzel, J. (2025). *Orthogonal Collocation on Finite Elements in Python*. [online] Medium. Available at: <https://medium.com/@jonas.weitzel/orthogonal-collocation-on-finite-elements-in-python-5bfb93f4a142> [Accessed 5 Dec. 2025].
- [13] Andersson, J.A.E., Gillis, J., Horn, G., Rawlings, J.B. and Diehl, M. (2018). CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1), pp.1–36. doi:<https://doi.org/10.1007/s12532-018-0139-4>.
- [14] Douglas, J. and Dupont, T. (1973). A finite element collocation method for quasilinear parabolic equations. *Mathematics of Computation*, [online] 27(121), pp.17–28. doi:<https://doi.org/10.1090/s0025-5718-1973-0339508-8>

Extending Liquid Phase Oligonucleotide Synthesis to a Two-Stage Membrane Cascade: Dynamic Modelling and Feasible Operating Region Analysis

Zoe Fang and Rachel Ha

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Liquid-phase oligonucleotide synthesis (LPOS) offers a promising alternative to solid-phase synthesis (SPOS), addressing concerns of high process mass intensity and low scalability. This work extends recent LPOS models and developments with branched homostar supports by implementing a two-stage membrane cascade for the diafiltration stages and comparing its performance against a single-stage configuration. A dynamic model was formulated through differential-algebraic equations and solved using numerical integration. Through a nested sampling feasibility analysis, we investigated the impact of key process parameters including deprotection time, coupling time, diafiltration time, recycle ratio, and second membrane volume, and analysed their effect on yield, purity, and residue targets. Results demonstrate that the two-stage system can achieve significantly higher yields. However, the system also enhances by-product retention, which accumulates building block impurities, causing a fundamental trade-off between targets. These findings provide clear operating guidelines for selecting a single-stage versus the two-stage membrane cascade design based on process requirements.

Key Words: Oligonucleotide, liquid-phase synthesis, organic solvent nanofiltration, feasibility analysis, membrane cascade

1. Introduction

Oligonucleotides are short, sequence-defined strands of ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) which are becoming more significant for modern medicine. Their high selectivity and specificity allow them to change gene expression, making them important in therapeutics and personalised medicine. This property allows them to be used in genetically driven or hard-to-treat diseases where conventional drugs often fall short. Beyond therapeutics, they are also used as primers in polymerase chain reaction (PCR) [1], and probes in gene sequencing [2]. They also play a clinical role in silencing or modifying disease-causing genes using antisense oligonucleotides (ASOs) and small interfering RNA (siRNA) [3].

Due to these diverse and high-impact applications, the oligonucleotide sector continues to expand rapidly, particularly in areas such as personalised medicine, biotechnology, and targeted therapeutics. The global oligonucleotide market is currently valued at \$12.5 billion in 2025, and is projected to grow to \$44.9 billion by 2033 – a 17.3% compound annual growth rate (CAGR) [4].

Currently, the industry-established production process for oligonucleotides is solid-phase synthesis (SPOS). The process involves an insoluble solid support (resin), to which nucleotide building blocks are added to form a growing oligonucleotide chain. However, SPOS presents significant economic and environmental drawbacks. Its scalability is limited by resin size, the process mass intensity (PMI) reaches 4,300 kg/kg of product and it has a poor atom economy e.g., 36 % for phosphoramidite-based synthesis [5] - the standard method for SPOS. These inefficiencies arise from the heterogeneous nature of the process, as excess reagents are required to ensure the reactions go to completion on the solid support, alongside large volumes of solvent (acetonitrile) which are needed for extensive washing after each cycle [6].

As an alternative, liquid-phase oligonucleotide synthesis (LPOS) is being researched to address these limitations. The process uses a soluble support for extension of the oligonucleotide chain with organic solvent nanofiltration (OSN). OSN utilises the size differences between products

and byproducts to achieve separation through a membrane and can be coupled with purification techniques such as precipitation [7]. LPOS enables better monitoring as continuous operation can maintain the product in solution, allowing for improved control over the oligonucleotide sequence via in-process liquid sampling [8].

Recent work by Saccardo and Chachuat [9] demonstrated a first dynamic model of LPOS with homostar supports, analysing the relationships between operating conditions and their effect on yield and purity. However, their analysis focused on a single-stage membrane configuration, which can require large volumes of fresh solvent, leading to compromising effects on yield after many cycles. Two-stage membrane cascade configurations have proved to increase yield and purity across pharmaceutical applications such as active pharmaceutical ingredient (API) purification [10], and peptide synthesis [11]. The cascade structure typically incorporates a recycle stream allowing extra retentate from the second membrane to be fed back to the reaction stage, retaining more product and unreacted reagents. This is economically important given the high cost of reagents and the expenses of waste disposal [12].

To ensure a more sustainable oligonucleotide production process, it is important to explore the addition of a two-stage membrane cascade to the current LPOS process and examine whether the feasible operating regions could be expanded. The objective of this paper is to extend the existing single-stage mathematical LPOS model to a two-stage membrane cascade and to investigate its potential for improving oligonucleotide yield and purity.

2. Background and Methodology

2.1 Process Description

Membrane-enhanced LPOS is comprised of three phases: loading of the soluble support; extension of the chain through iterative addition of building blocks to the homostar branches; and finally cleavage and purification. This work focused on the second phase, which operates through repeated cycles of four steps: deprotection, diafiltration, coupling, and a second diafiltration.

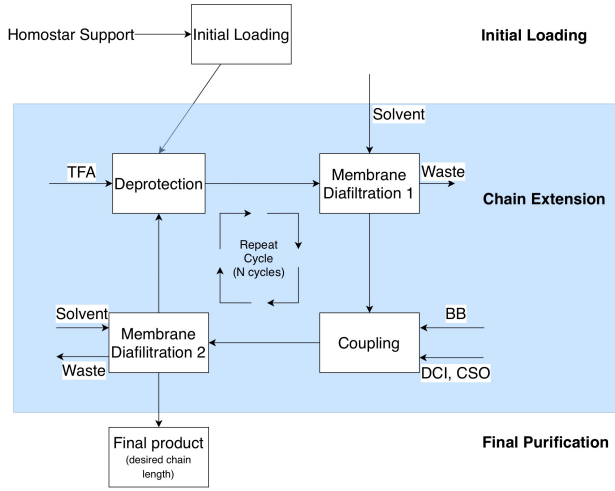


Figure 1: Overall process flowchart, adapted from Kim et al. [7]

In the deprotection step, the temporary protecting group dimethoxytrityl (Dmtr) on the N-terminus is removed using the catalyst, trifluoroacetic acid (TFA), generating the reactive site for the coupling reaction. This is followed by diafiltration, which is used to remove small species. An organic solvent is added continuously at the same rate as permeation to maintain a constant retentate volume.

In the coupling step, the nucleotide building block, dicyanoimidazole (DCI), and camphorsulfonyl oxaziridine (CSO) are added to react with the deprotected sites to extend the chains. Any remaining activated species are quenched by cyanoethanol to deactivate building blocks and prevent side reactions [13]. The second diafiltration is used to remove any unreacted building blocks, excess reagents, and quenching byproducts.

To further enhance the diafiltration, this work focuses on the implementation of a two-stage membrane diafiltration configuration, using the same OSN membrane for both stages. The retentate of each membrane therefore contains the same species, including all anchored oligonucleotides (deprotected, protected, and error species). The permeate of each membrane contains only small molecule reagents and by-products such as TFA, Dmtr fragments, unreacted building blocks, DCI, and CSO. The two retentate streams from each membrane stage are recombined before proceeding to the next reaction step, allowing all anchored oligonucleotides to remain together while improving small-molecule removal.

The required reagents are fed into the feed tank, and the mixture is circulated throughout the retentate loop to carry out the reaction. During diafiltration steps, the fresh solvent is taken from the solvent tank and added to the feed tank at the same rate as permeate being removed. The mixture is then passed sequentially through membrane 1 and membrane 2, where both membranes retain the large homostar anchored oligonucleotide. The two retentate streams exiting the membranes recombine and return to the feed tank before the next reaction step, ensuring that impurities are progressively washed out.

Accurate sequence formation during coupling is critical to ensure the correct biological function of the final oligonucleotide. When extending a target sequence such as A-B-C, several types of errors can occur such as truncated products (A-B) caused by incomplete coupling, deleted products where one nucleotide is skipped (A-C) and repeated units

(A-B-B) formed when residual building blocks react in later cycles. These impurities occur when there is insufficient reaction time or inadequate removal of small molecules such as DCI and CSO between the cycles.

LPOS has emerged as a solution to reduce the sequence error due to enabling control over reaction duration in coupling and diafiltration durations. The introduction of a two-membrane diafiltration cascade further strengthens this control.

2.2 Process Modelling

The mathematical model developed in this work built directly on the dynamic LPOS framework introduced by Saccardo and Chachuat [9]. We extended this framework to incorporate a two-membrane diafiltration cascade, assuming well-mixed retentate and permeate streams under constant-volume diafiltration.

For a target oligonucleotide composed of N building blocks, the process was represented as a sequence of N repeated cycles. During cycle $n = 1, \dots, N$, the newly added building block was represented by its concentration BB_n , while anchored oligonucleotides were described by their deprotected and protected states, OD_k and OP_k respectively. The variables were indexed by chain length and accounted for correctly extended sequences and error-containing species.

At the start of the process, anchored chains on each homostar branch contained no nucleotides.

$$OP_0(0) = HS \cdot NB \quad (1)$$

where HS (mol m^{-3}) represents the concentration of the homostar supports and NB represents the number of branches per support. The total concentration of all anchored oligonucleotides OT (mol m^{-3}) was expressed as:

$$OT(t) = \sum_{k=0}^N (OP_k(t) + OD_k(t)) \quad (2)$$

The LPOS cycle alternated between reaction and membrane-separation steps, each governed by different dynamic equations. The overall system was formulated as a multistage dynamic model in which state discontinuities may arise at the transitions between consecutive process stages.

2.2.1 Deprotection

During the deprotection stage of cycle n , all protected anchored chains with lengths $k = 0, \dots, n-1$ were converted into their deprotected forms driven by the TFA catalyst.

$$\begin{aligned} \dot{OD}_k(t) &= -\dot{OP}_k(t) \\ &= k_D OP_k(t) TFA(t), \quad k = 0, \dots, n-1 \end{aligned} \quad (3)$$

where k_D denotes the kinetic constant for deprotection.

At the start of the deprotection stage, TFA (mol m^{-3}) was introduced to the mixture through an instantaneous pulse addition.

$$TFA(t) = TFA(t_{D,n}^+) = TFA(t_{D,n}^-) + TFA_{in} \quad (4)$$

Once added, TFA remained constant throughout the duration of the deprotection stage as the catalyst was regenerated. The other species in the mixture remained unchanged, as they were not involved in the deprotection reaction.

2.2.2 Single-Stage Diafiltration

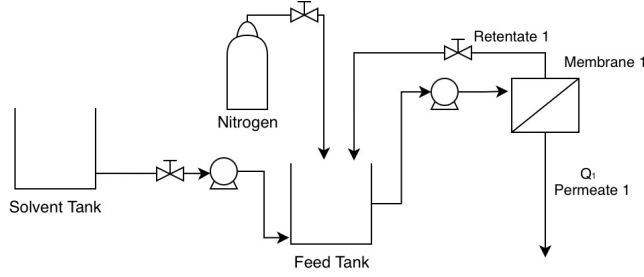


Figure 2: LPOS Process Layout for Single-Membrane, adapted from Kim et al. [14]

The single-membrane operated as described in Saccardo and Chachuat's [9] original one membrane LPOS model.

Its dynamics were governed by a constant-rejection washout equation:

$$\dot{X}(t) = -\frac{Q_1}{V_1} (1 - r_X) X(t) \quad (5)$$

$$\forall X \in \{OP_k, OD_k, TFA, BB_n, DCI, CSO\}$$

where V_1 (m^3) denotes the constant liquid volume on the retentate side of membrane 1, and r_X is the observed membrane rejection of solute X , defined as the fraction of solute retained by the membrane (nominal values are shown in Table 1).

The volumetric flux Q_1 ($\text{m}^3 \text{s}^{-1}$) is defined by:

$$Q_1 = AK\Delta p_1 \quad (6)$$

where A (m^2) is the membrane area, K ($\text{m s}^{-1} \text{bar}^{-1}$) is the permeability of the membrane and Δp_1 (bar) is the transmembrane pressure difference.

2.2.3 Two-Stage Diafiltration

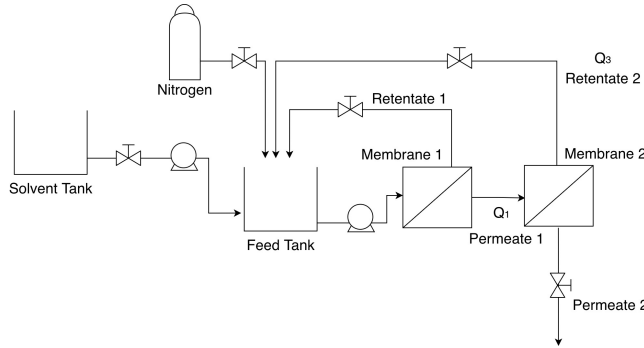


Figure 3: LPOS Process Layout, adapted from Kim et al. [14]

The extension arose from the introduction of the second membrane, which follows the cascade structure of Kim et al. [14]. The two-membrane diafiltration was modelled as a single continuous operation, rather than two discrete filtration steps. This configuration introduces an additional set of dynamic equations to describe solute concentration in the second stage, with the permeate from the first membrane as the inlet to the second. The volume of the second membrane on the retentate side was denoted as V_2 . Solvent addition was applied at a constant rate that matched the permeate outflow,

ensuring that the retentate volumes V_1 and V_2 remained constant. The concentrations of solute X in the first and second membrane stages, denoted as $X_a(t)$ and $X_b(t)$ respectively, were governed by:

$$\dot{X}_a(t) = \left(\frac{1}{V_1}\right) [-Q_1 X_a(1 - r_X) + Q_3 X_b] \quad (7)$$

$$\dot{X}_b(t) = \left(\frac{1}{V_2}\right) [Q_1 X_a(1 - r_X) - Q_2 X_b(1 - r_X) - Q_3 X_b] \quad (8)$$

where $Q_1 X_a(1 - r_X)$ represents the solute that permeated through the first membrane and entered the second stage, and $Q_3 X_b$ represents the recycle stream from the second stage.

At the end of the diafiltration period, the retentate material from both membranes was recombined to form the initial state of the subsequent step. For the first diafiltration in cycle 1, the second stage retentate initially contained no material, therefore giving the initial condition:

$$X_b(0) = 0 \quad (9)$$

Both membrane stages used the same OSN membrane as described in Saccardo and Chachuat's original single-membrane LPOS model. The membrane area A and permeability K were therefore identical for both membranes. Since the transport characteristics of the membrane remained unchanged, the rejection coefficients for each species were the same for both membranes. As a result, differences in the diafiltration performance arose from the flow configuration, values of the recycle ratio, and volume used in the second membrane and the overall total diafiltration constant.

$$Q_1 = AK\Delta p_1 \quad (10)$$

$$r = \frac{Q_3}{Q_1} = \frac{AK\Delta p_1 - AK\Delta p_2}{AK\Delta p_1} \quad (11)$$

$$Q_2 = Q_1 - Q_3 \quad (12)$$

The second membrane did not require an explicit pressure difference equation as the operating pressure was determined by the recycle ratio which can be calculated from Equation (11). The recycle ratio specified the fraction of permeate from the first membrane that was recirculated into the main feed tank: Q_3 denoted the recycle stream flowrate, whilst the remainder Q_2 exited the system as permeate waste. The recycle ratio r was treated as an independent variable, controlled via valves determining the permeate stream from the second membrane.

The retentate volume of the first membrane, V_1 , was fixed at the value from previous literature [9], whilst the second membrane volume, V_2 , was introduced as an independent variable which could be adjusted to reflect different operating conditions.

2.2.4 Coupling

The coupling stage of cycle n described the attachment of n th building block to the deprotected anchored chains present in the mixture.

Table 1: Design and model parameter values adapted from [9]

Parameter	Nominal value
HS	5.0 mol m^{-3}
NB	4
k_D	$3.6 \times 10^{-5} \text{ m}^3 \text{ mol}^{-1} \text{ s}^{-1}$
k_C	$2.7 \times 10^{-8} \text{ m}^9 \text{ mol}^{-3} \text{ s}^{-1}$
k_S	$4.3 \times 10^{-6} \text{ m}^3 \text{ mol}^{-1} \text{ s}^{-1}$
BB^{in}	$1.5 \text{ eq}^\dagger$
TFA^{in}	$8.0 \text{ eq}^\dagger$
DCI^{in}	$4.0 \text{ eq}^\dagger$
CSO^{in}	$6.0 \text{ eq}^\dagger$
A	$8.5 \times 10^{-2} \text{ m}^2$
V_1	$9 \times 10^{-4} \text{ m}^3$
Δp_1	8.9 bar
K	$5.6 \times 10^{-7} \text{ m s}^{-1} \text{ bar}^{-1}$
r_{OP}	0.998
r_{OD}	0.998
r_{BB}	0.40
r_{TFA}	0.33
r_{DCI}	0.33
r_{CSO}	0.33

† Equivalent amounts relative to the initial anchored oligonucleotides
 $OT(0) = HS \cdot NB$.

The step transformed the deprotected species of length $k = 0, \dots, n-1$ into their protected counterparts of length $k+1$, using the building block, DCI and CSO. The overall stoichiometric equation was represented as:



where $k = 0$ to $n-1$ and $n = 1$ to 8 to form octamers. The case where $k = n-1$ corresponded to the desired synthesis pathway, forming the correctly extended sequence. When $k < n-1$, the coupling reaction extended the previously deleted or truncated chains. Repeated-unit impurities were not included in our model as residual building blocks from early cycles were assumed to be fully quenched and deemed negligible due to extremely low concentrations.

The mass balances for the protected and deprotected oligonucleotides were denoted as follows:

$$\begin{aligned} \dot{OP}_{k+1}(t) &= -\dot{OD}_k(t) \\ &= k_C BB_n(t) OD_k(t) DCI(t) CSO(t) \end{aligned} \quad (14)$$

where k_C and k_S denote the kinetic constants for the coupling and side reactions respectively, and BB_n (mol m^{-3}), DCI (mol m^{-3}), and CSO (mol m^{-3}) represent their respective concentrations. The building blocks and CSO were also involved in the side reaction, so their material balances followed:

$$\begin{aligned} \dot{BB}_n(t) &= \dot{CSO}(t) \\ &= -k_C BB_n(t) \sum_{k=0}^{n-1} OD_k(t) DCI(t) CSO(t) \\ &\quad - k_S BB_n(t) CSO(t) \end{aligned} \quad (15)$$

$$\dot{DCI}(t) = -k_C BB_n(t) \sum_{k=0}^{n-1} OD_k(t) DCI(t) CSO(t) \quad (16)$$

Pulse additions of building blocks BB^{in} and reactants DCI^{in} , CSO^{in} (mol m^{-3}) were made at start time $t_{C,n}^-$ and denoted as follows:

$$BB_n(t_{C,n}^+) = BB_n(t_{C,n}^-) + BB^{\text{in}} \quad (17)$$

$$DCI(t_{C,n}^+) = DCI(t_{C,n}^-) + DCI^{\text{in}} \quad (18)$$

$$CSO(t_{C,n}^+) = CSO(t_{C,n}^-) + CSO^{\text{in}} \quad (19)$$

The concentrations of all other species in the mixture not participating in the coupling reaction remained constant.

2.2.5 Key Performance Indicators

To evaluate how process performance changes when moving from a single-membrane to a two-stage membrane cascade, three metrics were quantified that highlight the differences in product formation, selectivity, and removal efficiency across the full synthesis sequence.

The overall yield reflected how much correctly assembled fully protected oligonucleotide after N cycles to the theoretical maximum.

$$Y = \frac{OP_N(t_{F2,N}^-)}{HS \cdot NB} \quad (20)$$

The product purity was the proportion of correctly sequenced oligonucleotides to all anchored species present after the final coupling step.

$$P = \frac{OP_N(t_{F2,N}^-)}{OT(t_{F2,N}^-)} \quad (21)$$

Keeping purity high was important because incomplete coupling or the formation of sequences containing errors added to the impurities, compromising experimental reliability. Additionally, impurities and correctly sequenced products possesses similar physical properties, making downstream separation particularly challenging.

To monitor the accumulation of unreacted building blocks, the ratio of remaining building blocks after the first diafiltration to the total anchored material was tracked for each cycle. The worst (largest) ratio was defined as:

$$R = \max_{n=1, \dots, N} \left[\frac{\sum_{k=1}^n BB_k(t_{C,n}^-)}{OT(t_{C,n}^-)} \right] \quad (22)$$

Maintaining a low value of R limited the possibility of unintended reactions in later cycles.

3. Feasibility

The dynamic model implemented the mass-balance equations for the four LPOS stages simulated over $N = 8$ using CRONOS¹, a C++ solver interfaced with Python. All equations were compiled into an expression tree and represented

¹<https://github.com/omega-icl/cronos>

as a directed acyclic graph (DAG) using MC++², so that CRONOS could evaluate it directly during feasibility sampling.

Feasibility analysis was carried out using MAGNUS³, which performed nested sampling. The algorithm operated by producing a population of live points within the operating regions and updating this population through likelihood-guided exploration. At each iteration, new candidate points were generated by MAGNUS, and the dynamic model evaluated via CRONOS. A likelihood value was then computed reflecting how well the resulting performance indicators satisfied the feasibility specifications. A candidate replaced the lowest-likelihood live point whenever it performed better, whilst the discarded point was stored for reconstructing the feasible region and classified as dead. Kinetic, membrane, and concentration parameter values were obtained from Saccardo and Chachuat [9], and are listed in Table 1.

3.1 Single-stage membrane diafiltration

To ensure direct comparison with the feasibility regions reported by Saccardo and Chachuat [9], the same four operational decisions were varied: deprotection duration, Δt_D , coupling duration, Δt_C , first diafiltration duration, Δt_{F1} , and second diafiltration duration, Δt_{F2} . Likewise, the same prior bounds as the original LPOS analysis were retained:

$$\begin{aligned} 5 \leq \Delta t_D \leq 20 \text{ min}, & \quad 80 \leq \Delta t_C \leq 120 \text{ min}, \\ 0 \leq \Delta t_{F1} \leq 80 \text{ min}, & \quad 0 \leq \Delta t_{F2} \leq 80 \text{ min} \end{aligned}$$

For the single-membrane configuration, nested sampling was performed using 15,000 live points and 4,500 candidate proposals to generate sufficient feasible points for reliable convex hull analysis.

3.2 Two-stage membrane cascade diafiltration

Two additional degrees of freedom were introduced in the two-stage diafiltration cascade: volume on the retentate side of the second membrane (V_2), and the recycle ratio (r), increasing the number of dynamic states and couplings in the model. Given that $V_1 = 9 \times 10^{-4} \text{ m}^3$, to allow for a pressure difference between the two membranes, the bounds on the new variables were set as follows:

$$0 \leq V_2 \leq 9 \times 10^{-4} \text{ m}^3, \quad 0 \leq r \leq 1$$

Using 15,000 live points and 4,500 proposals would have driven the sampler to concentrate rapidly around the densest feasible region, so each trellis subplot in Figures 4 to 6 used a reduced population of 5,000 live points and 1,500 proposals to ensure broader exploration.

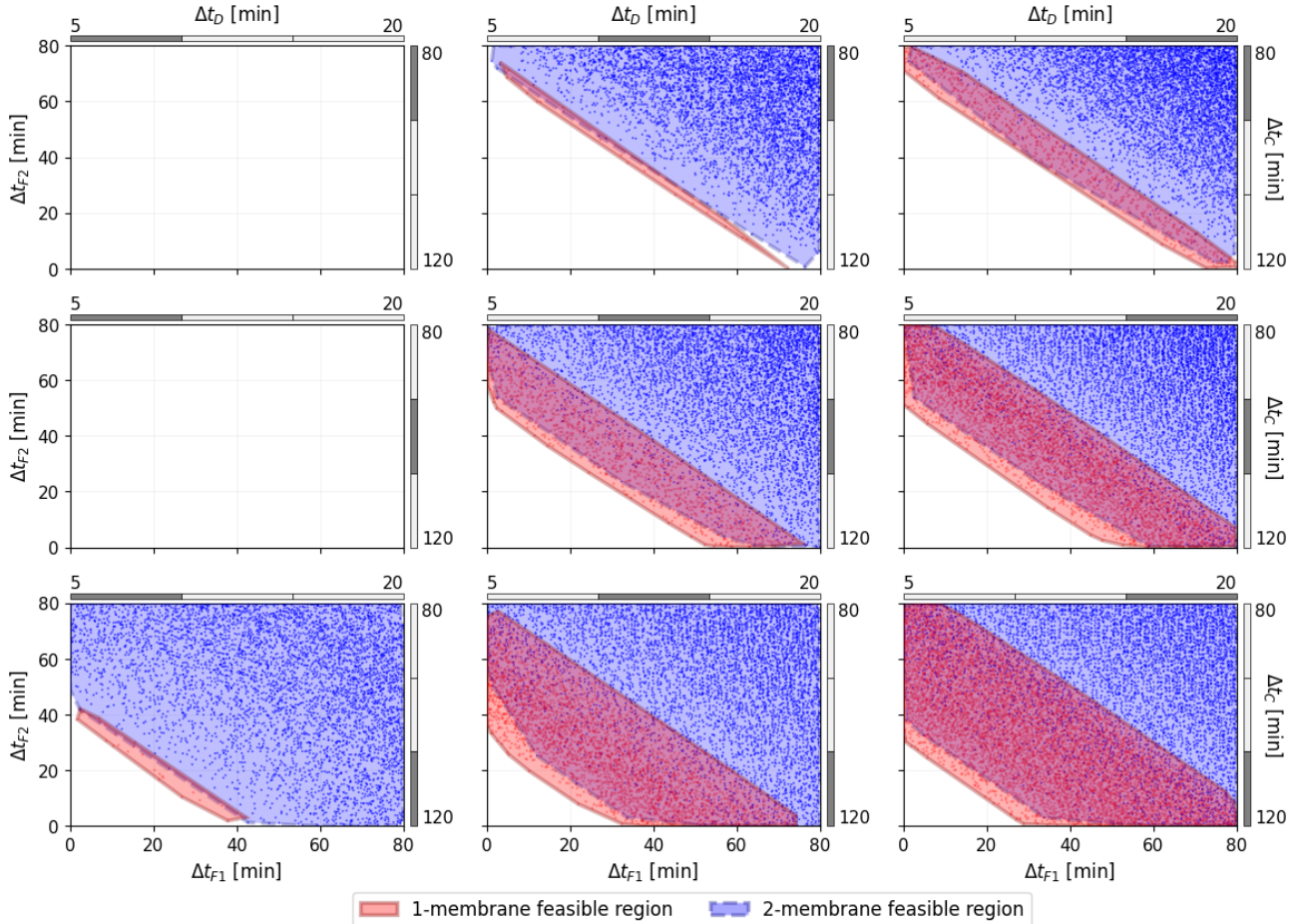


Figure 4: Trellis chart comparing the feasible regions of single-membrane and two-membrane at constraints of $Y \geq 0.9$, $P \geq 0.9$, $R \leq 0.01$

²<https://github.com/omega-icl/mcpp>

³<https://github.com/omega-icl/magnus>

For the two-membrane cascade analyses in Section 4.2, a moderate setting of 10,000 live points and 3,000 proposals was sufficient. The experiments ran were able to achieve stable convergence of the nested-sampling algorithm, allowing the key interactions governing the feasibility to be captured.

4. Results and Discussion

4.1 Initial Feasible Operating Set Analysis

The first stage of the analysis considered the nominal feasibility thresholds to achieve high yield and purity, and low building block residue, matching Saccardo and Chachuat [9]:

$$Y \geq 0.9, \quad P \geq 0.9, \quad R \leq 0.01$$

Using this criteria, the feasible operating regions for both the single-membrane and two-stage membrane cascade were evaluated under identical sampling bounds listed in Section 3.1 across all process stage durations.

Figures 4 to 6 show trellis plots of the four-dimensional feasible region for the single and two-stage membrane configurations from MAGNUS. Subplot axes show diafiltration times Δt_{F1} and Δt_{F2} , with each row and column corresponding to combinations of different tertile ranges of Δt_D and Δt_C , as indicated by the shaded grey bars.

Figure 4 shows regions of feasible points for the single-membrane and two-stage membrane cascade under all three constraints simultaneously. The single-stage model shows no feasible points at low Δt_D and Δt_C , with a feasibility band appearing with increasing width as coupling and deprotection times are extended. Short coupling and deprotection times lead to lower product conversion and higher residual

building block concentration, restricting feasibility. Conversely, the two-stage system (blue convex hulls) expands across the full upper triangular region in all subplots. Where the single-stage system fails to retain sufficient product, the recycle stream of the two-stage membrane cascade recovers the oligonucleotide from the permeate of the first membrane, thus achieving the target yield. The lower diagonal feasible boundaries indicate the minimum diafiltration times required to eliminate building block residue. The single-stage system exhibits upper boundaries where excessive combined diafiltration times cause oligonucleotide loss, creating the shape of the feasible band. The two-stage cascade is free from this upper boundary within these bounds, as the recycle stream prevents these losses.

To investigate the effect of individual constraints, experiments imposing only $Y \geq 0.9$ (Figure 5) and $R \leq 0.01$ (Figure 6) were conducted. When isolating the $Y \geq 0.9$ constraint, the single-stage membrane configuration shows a feasible region bounded by the sum of diafiltration times, confirming that the yield imposes the upper boundary as shown in Figure 4, whereas the two-stage system demonstrates feasibility across all combinations of Δt_{F1} and Δt_{F2} . In Figure 5, the single-stage shows that increasing diafiltration times leads to greater loss of oligonucleotides across the membrane, imposing an upper bound on Δt_{F1} and Δt_{F2} . As the deprotection time increases across the rows, the upper bound relaxes, expanding the area of the lower red triangles which extend to longer diafiltration durations. This suggests that Δt_D is the dominant parameter in determining the single-stage yield.

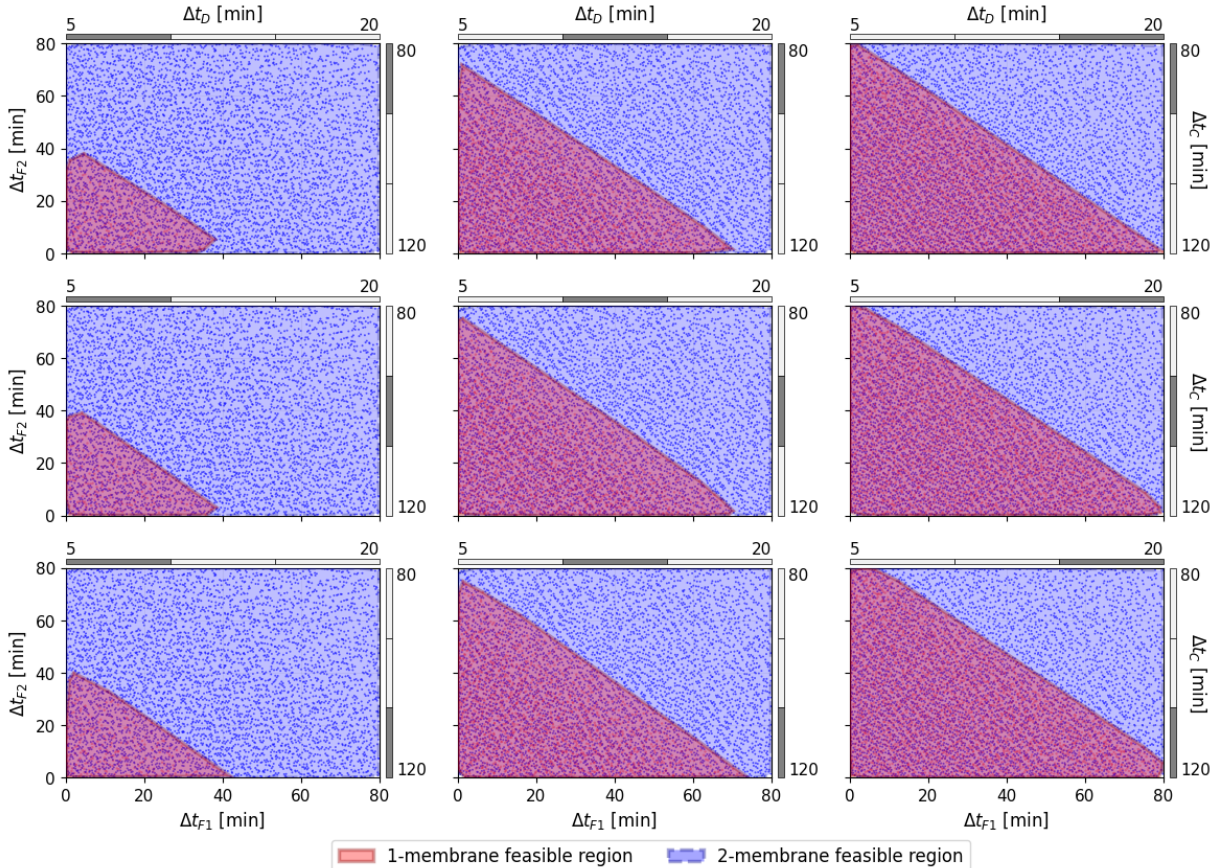


Figure 5: Trellis chart comparing the feasible region of single-membrane and two-membrane at a constraint of $Y \geq 0.90$

In contrast, the two-membrane cascade consistently satisfies the yield target due to a larger recovery from the recycle stream, allowing for more combinations of diafiltration times to remain feasible. The greater blue area of the feasible region for the two-stage membrane demonstrates that yield is less sensitive to Δt_D in this configuration, implying that the yield may be upper bounded by diafiltration times outside the experimented range, suggesting potential for further exploration.

In contrast to yield behaviour, the individual residue constraint $R \leq 0.01$ on the two-membrane cascade system exhibits a smaller feasible region than the single-stage system. The upper triangular shape in Figure 6 indicates that longer combined diafiltration times are required to achieve adequate removal of residual building blocks, exhibiting a lower bound. This behaviour in the two-stage (blue) region arises from the same recycle stream which enhances yield: while recovering the desired product, it simultaneously returns unreacted building blocks to the coupling stage in following cycles, leading to impurity accumulation that requires extended diafiltration for removal.

Notably, the feasible regions expand with longer coupling times (down columns), indicating that extended coupling improves the management of residue across both systems. Δt_C significantly influences the size of the feasible operating regions as it allows for more complete building block BB_n reactions, preventing sequence errors of oligonucleotides. This longer coupling time enables shorter diafiltration times with $\Delta t_{F1} + \Delta t_{F2}$ decreasing from 80 minutes at low Δt_C to 40 minutes at high Δt_C . The relationship between reaction and separation times reveals a compromise is needed

if wanting to minimise total operation time. Overall, the two-stage cascade offers a more flexible operating space, reflecting its capacity to meet the required constraints over a wider range of conditions. While the two-stage system offers higher yield performance (Figure 5), the residue constraint presents the greater challenge, highlighting the importance of coupling time and its effect on feasibility. The feasibility regions in Figure 4 for the single-stage system are determined by the intersection of the feasible regions for yield and residue. However, it is important to note this is not the case for the two-stage system due to the variation in additional parameters, which expand the feasible space. A fundamental trade-off between yield and residue removal is therefore exhibited in the two-stage membrane cascade.

4.2 Additional Feasible Operating Set Analysis

4.2.1 Six-Parameter Feasible Region Analysis

Following the improved feasibility results at the lower constraints evaluated, the KPIs were changed to evaluate the performance of the two-stage system at higher yield and purity to determine its capability at higher desired targets:

$$Y \geq 0.97, \quad P \geq 0.97, \quad R \leq 0.01$$

Figure 7 shows a corner plot of the feasible region of operating parameters for a two-membrane cascade for LPOS. Each diagonal panel shows the marginal distribution for one parameter (Δt_D , Δt_{F1} , Δt_C , Δt_{F2} , V_2 , r), and the scatter plots display the two-dimensional projections of feasible operating conditions (blue scatter). Operating ranges for all parameters were maintained as previously specified in Section 4.1.

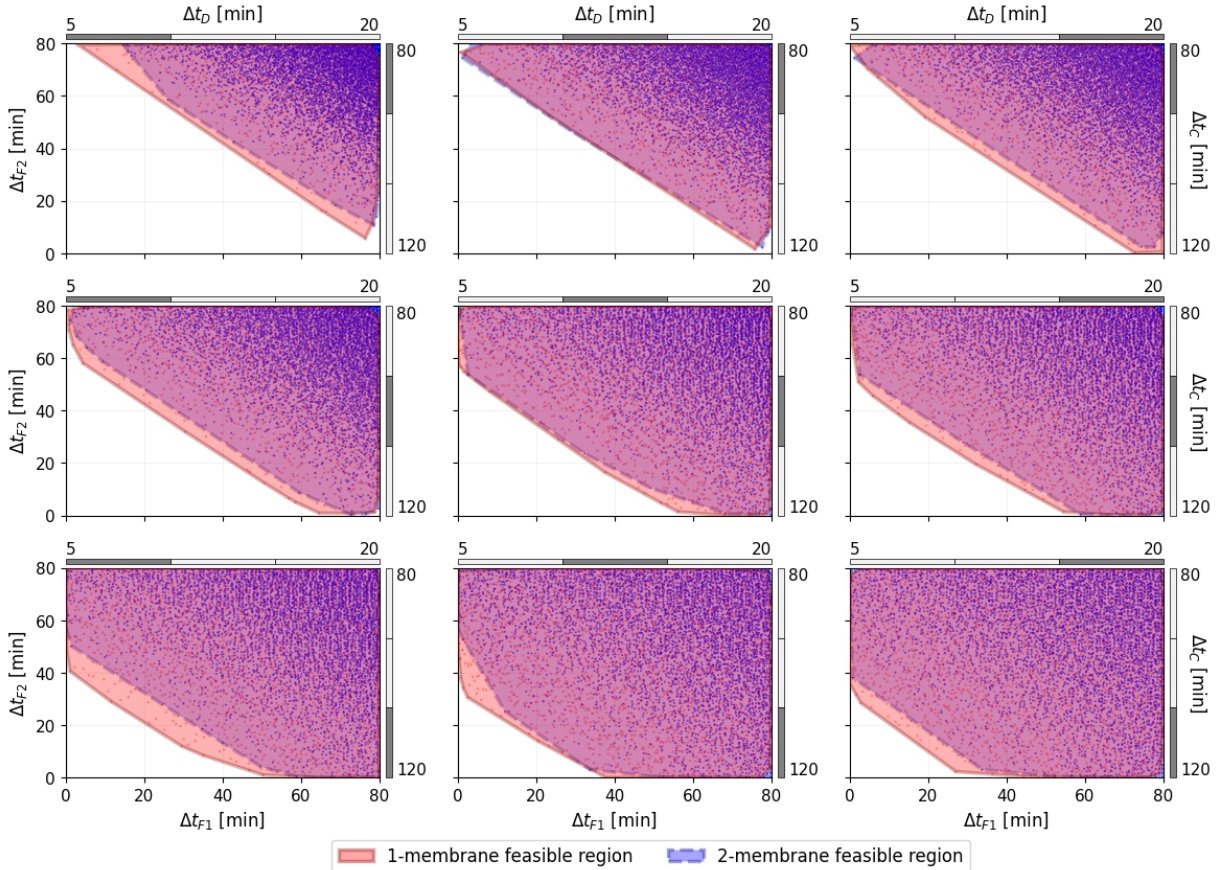


Figure 6: Trellis chart comparing the feasible region of the single-membrane and two-stage membrane cascade at a constraint of $R \leq 0.01$

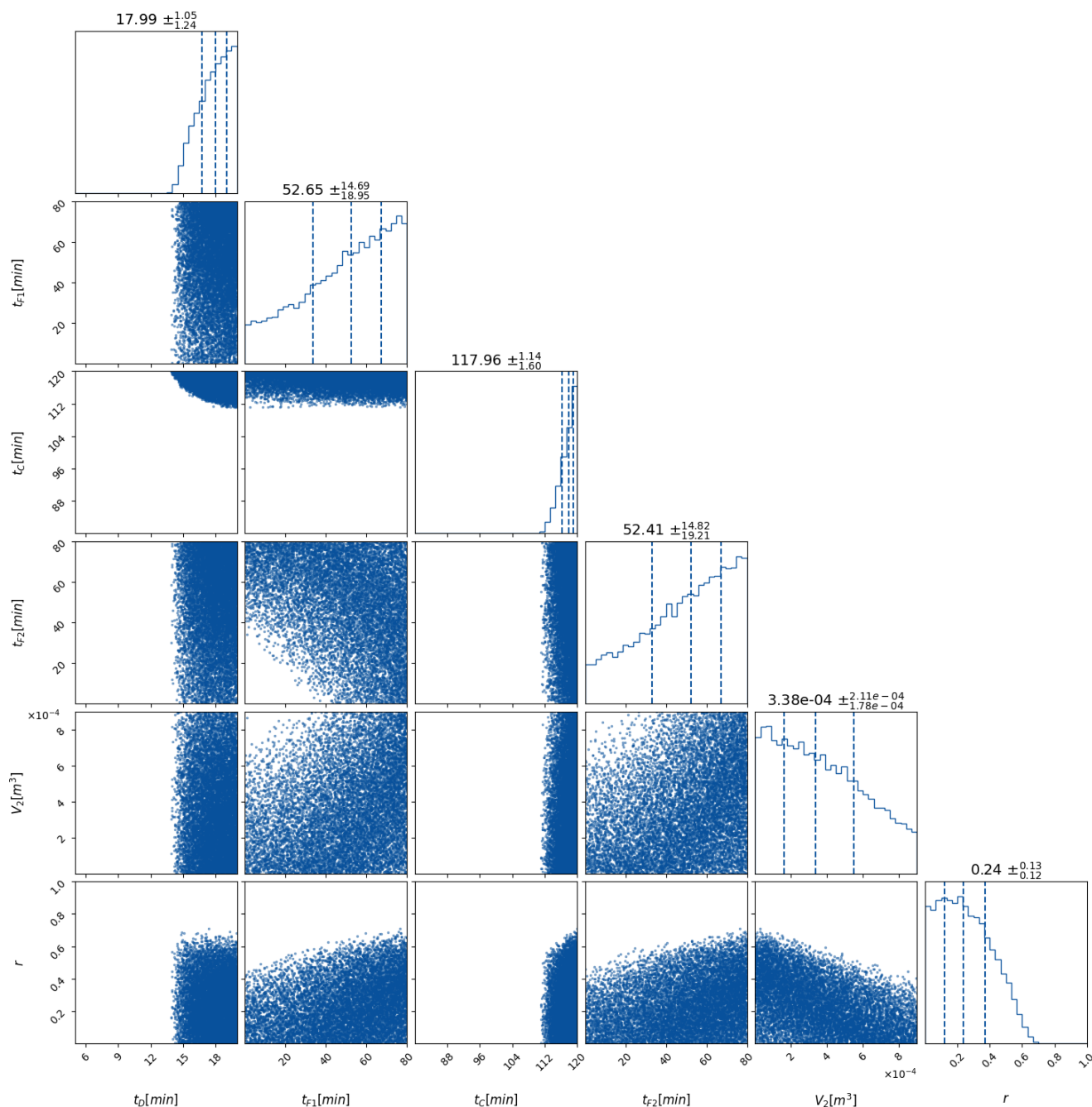


Figure 7: Corner plot of feasible operating parameters for a two-stage membrane cascade with constraints $Y \geq 0.97$, $P \geq 0.97$, $R \leq 0.01$. Diagonal panels show the histograms of the parameter with vertical dashed lines showing their interquartile values.

In order to reach yield and purity ≥ 0.97 , minimum deprotection and coupling times of 14 min and 111 min respectively are required from Figure 7. Insufficient times for deprotection and coupling result in a higher possibility of truncated or incorrectly sequenced oligonucleotides, and the high density of points leading to well-defined boundaries indicates hard constraints.

The projection of $t_{F1} - t_{F2}$ shows the absence of feasible points in the lower-left corner, indicating that the minimum total diafiltration time, $(\Delta t_{F1} + \Delta t_{F2})$, of 40 minutes is required to eliminate building block residue. The weak boundaries shown by sparse scatter points suggest areas of constraint overlap, suggesting there is greater operational

flexibility for those parameters.

The projection of $V_2 - r$ demonstrates an inverse relationship, as a higher membrane volume in the second stage requires lower recycle ratios, with the median value of r at 0.24. Recycle ratios of 0.75 and above prove infeasible, as fewer by-products are expelled from the system when a higher retentate flow is recycled back to the feed tank. The median value of V_2 stands at $3.38 \times 10^{-4} \text{ m}^3$ (approximately one-third of V_1). This arises from reduced molar quantities of compounds from the first membrane permeate, requiring a smaller diafiltration volume to maintain the concentration gradient driving membrane separation. The projections of V_2 paired with t_{F1} and t_{F2} each display linear feasible bound-

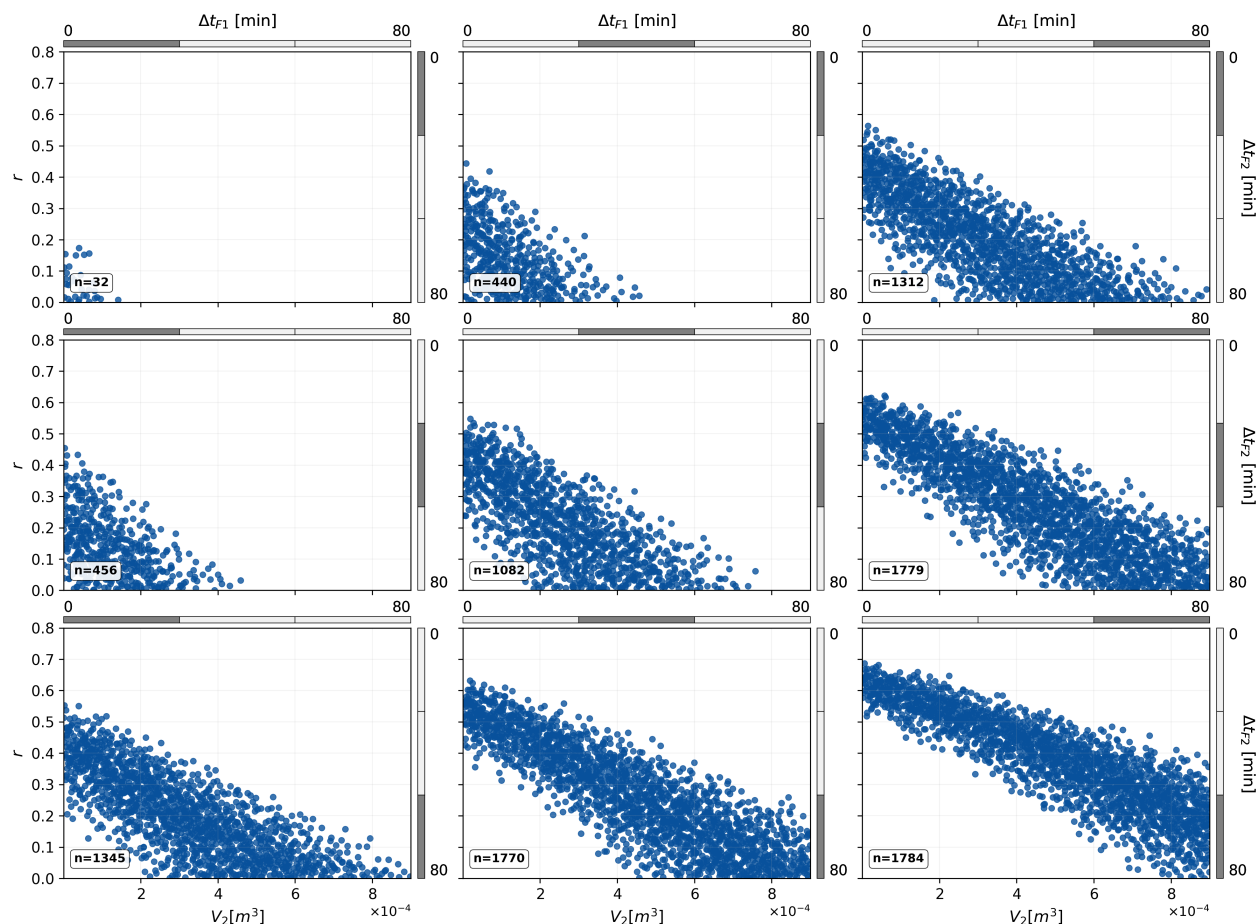


Figure 8: Two-membrane slices with fixed $\Delta t_D = 18$ min, $\Delta t_C = 118$ min with operation constraints at $Y \geq 0.97$, $P \geq 0.97$, $R \leq 0.01$

aries forming positively diagonal feasible regions, as a larger V_2 requires a longer diafiltration time to achieve the same degree of washout.

4.2.2 Sensitivity Analysis of V_2 and r

Following the initial exploration of six parameters, the deprotection and coupling times were fixed at their median values of 18 and 118 minutes respectively, to conduct a more rigorous analysis of the new design parameters V_2 and r . Figure 8 shows the trellis plot for the four-dimensional feasible region with V_2 and r on the subplot axes at different ranges of Δt_{F1} and Δt_{F2} .

Operating ranges of recycle ratio and volume increase for longer diafiltration times, as reflected by the increasing number of feasible points from 32 in the upper-left subplot to 1784 in the lower-right, indicating increased operational flexibility. The negatively sloped linear relationship between V_2 and r shown by the feasible bands is formed by upper and lower bound constraints: yield and residue respectively. An approximate symmetry is shown about the diagonal, with similar feasible point counts and band shapes (e.g., 440 versus 456 and 1312 versus 1345), suggesting that Δt_{F1} and Δt_{F2} contribute equally to meeting performance targets. Physically, higher recycle ratios ensure that more material is returned to the first stage providing additional volume and removal capacity, which reduces the second membrane volume

required to achieve the yield target. However, increasing recycle ratio requires longer diafiltration times, as shown by the upward shift of the feasible region across rows and down columns. When extending diafiltration times, the width of the feasible band increases and its gradient decreases, indicating a reduced sensitivity of volume to recycle ratio. For extended diafiltration times, higher volumes are needed to reach the yield target, which increases solvent consumption and reduces sustainability. Low recycle ratios and volumes are suggested for balancing sustainability whilst still meeting constraints, therefore operation in low to moderate diafiltration times is recommended. Specifically $\Delta t_{F1} \leq 53$ min, $\Delta t_{F2} \leq 53$ min, $r \leq 0.45$ and $V_2 \leq 4.5 \times 10^{-4}$ m³. These guidelines follow from the subplots in the upper-middle row and first-middle column.

5. Conclusions

This work extends the single-membrane dynamic LPOS model to a two-membrane cascade and analyses the feasible operating space through nested sampling. Advantages of the two-membrane cascade include enlarging the feasible operating region for achieving higher yield and purity, retaining unreacted reagents, and providing greater operational flexibility regarding process timings. However, larger retention of unreacted building blocks introduces a fundamental trade-off due to the recycle ratio and tightens the residue con-

straints, therefore requiring longer diafiltration times. These findings align with experimental observations in peptide synthesis using membrane cascades [15]. Overall, this study shows the interplay between how membrane configuration affects LPOS performance and provides a threshold for selecting and optimising this process.

Future developments of this work should focus on the applicability of the two-membrane cascade LPOS model. Obtaining experimental values of the kinetic constants and membrane rejection values would be the natural next step as this would improve model accuracy, since the current model relies on nominal literature parameters. Rejection values could be obtained experimentally and modelled as non-constant values. The exploration of different membrane types and mixed-selectivity cascade designs is an important direction for future work. The current work assumes identical membranes in both cascade stages. OSN literature shows that membrane selectivity strongly influences performance [16]. Tailored selectivity could be introduced, for example, a higher rejection of building blocks could alleviate the residue limitations observed in the membrane cascade whilst still preserving high yield. Yield and purity KPIs can further be varied independently; however, maintaining both at high levels remains essential for this process, which could be supported by varying the durations of each stage across the eight cycles rather than applying uniform timing.

6. Acknowledgements

The authors would like to thank Alberto Saccardo for his guidance and invaluable input.

References

- [1] N. Technologies. "What do oligonucleotides do in pcr?" (2025), [Online]. Available: <https://nanosttring.com/blog/what-do-oligonucleotides-do-in-pcr/> (visited on 10/12/2025).
- [2] M. Hendling and I. Barišić, "In-silico design of DNA oligonucleotides: Challenges and approaches," *Comput. Struct. Biotechnol. J.*, vol. 17, pp. 1056–1065, Jul. 2019, PMC6700205. doi: 10.1016/j.csbj.2019.07.008.
- [3] Exactmer. "The growing demand for oligonucleotide synthesis: Uses, challenges, and future opportunities." (2025), [Online]. Available: <https://exactmer.com/the-growing-demand-for-oligonucleotide-synthesis-uses-challenges-and-future-opportunities/> (visited on 11/09/2025).
- [4] Market.us. "Oligonucleotide synthesis market report." (2024), [Online]. Available: <https://market.us/report/oligonucleotide-synthesis-market/> (visited on 08/12/2025).
- [5] Y. S. Sanghvi, L. Ferrazzano, W. Cabri, and A. Tolomelli, "Sustainable approaches in solid-phase oligonucleotide synthesis: Current status and future directions," in *Sustainability in Tides Chemistry: Green Approaches*, A. Tolomelli, L. Ferrazzano, and W. Cabri, Eds., Cambridge, UK: RSC, 2024, ch. 10. doi: 10.1039/9781837674541.
- [6] Y. Xu, "Spos and lpos: Complementary or competitive processes?" SynerG Biopharma, Tech. Rep., 2025. [Online]. Available: <https://synergbiopharma.com/white-paper/solid-phase-oligonucleotide-synthesis-spos-and-liquid-phase-oligonucleotide-synthesis-lpos-complementary-or-competitive-processes/>.
- [7] J. F. Kim, P. R. J. Gaffney, I. B. Valtcheva, G. Williams, A. M. Buswell, M. S. Anson, and A. G. Livingston, "Organic solvent nanofiltration (OSN): A new technology platform for liquid-phase oligonucleotide synthesis (LPOS)," *Org. Process Res. Dev.*, vol. 20, no. 8, pp. 1439–1452, 2016. doi: 10.1021/acs.oprd.6b00139.
- [8] P. R. J. Gaffney, J. F. Kim, I. B. Valtcheva, G. D. Williams, M. S. Anson, A. M. Buswell, and A. G. Livingston, "Liquid-phase synthesis of 2'-methyl-rna on a homostar support through organic-solvent nanofiltration," *Chem. Eur. J.*, vol. 21, no. 26, pp. 9535–9543, 2015. doi: 10.1002/chem.201501001.
- [9] A. Saccardo and B. Chachaut, "Model-based analysis of membrane-enhanced liquid-phase oligonucleotide synthesis," *IFAC-PapersOnLine*, vol. 59, no. 6, pp. 277–282, 2025. doi: 10.1016/j.ifacol.2025.07.158.
- [10] L. Peeva, J. d. S. Bursal, I. Valtcheva, and A. G. Livingston, "Continuous purification of active pharmaceutical ingredients using multistage organic solvent nanofiltration membrane cascade," *Chem. Eng. Sci.*, vol. 116, pp. 183–194, 2014. doi: 10.1016/j.ces.2014.04.022.
- [11] S. So, L. G. Peeva, E. W. Tate, R. J. Leatherbarrow, and A. G. Livingston, "Organic solvent nanofiltration: A new paradigm in peptide synthesis," *Organic Process Research & Development*, vol. 14, pp. 1313–1325, 2010. doi: 10.1021/op100156y.
- [12] B. I. Andrews, F. D. Antia, S. B. Brueggemeier, L. J. Diorazio, S. G. Koenig, M. E. Kopach, H. Lee, M. Olbrich, and A. L. Watson, "Sustainability challenges and opportunities in oligonucleotide manufacturing," *J. Org. Chem.*, vol. 86, no. 1, pp. 49–61, 2021. doi: 10.1021/acs.joc.0c02291.
- [13] A. H. El-Sagheer and H. Lönnberg, "Solution-phase synthesis of oligonucleotides on a soluble support," *Beilstein J. Org. Chem.*, vol. 13, pp. 1342–1357, 2017. doi: 10.3762/bjoc.13.134.
- [14] J. F. Kim, A. M. F. da Silva, I. B. Valtcheva, and A. G. Livingston, "When the membrane is not enough: A simplified membrane cascade using organic solvent nanofiltration (osn)," *Sep. Purif. Technol.*, vol. 116, pp. 277–286, 2013. doi: 10.1016/j.ifacol.2025.07.158.
- [15] W. Chen, M. Sharifzadeh, N. Shah, and A. G. Livingston, "Iterative peptide synthesis in membrane cascades: Untangling operational decisions," *Computers and Chemical Engineering*, vol. 115, pp. 275–285, Jun. 2018, ISSN: 0098-1354. doi: 10.1016/j.compchemeng.2018.04.024.
- [16] P. Marchetti, L. Peeva, and A. Livingston, "The selectivity challenge in organic solvent nanofiltration: Membrane and process solutions," *Annu. Rev. Chem. Biomol. Eng.*, vol. 8, pp. 473–497, 2017. doi: 10.1146/annurev-chembioeng-060816-101325.

Techno-economic Analysis of Hybrid Energy Systems for Powering Hyper-scale Data-centres in the UK

Joseph Ponsford¹

¹*Imperial College London.*

Abstract

The rapid expansion of Artificial Intelligence (AI) is driving a surge in energy demand for hyper-scale data centres, projected to triple the UK's sector consumption by 2035. This growth coincides with significant grid constraints in the UK, where £393 million was spent in 2024 to curtail renewable wind generation, predominantly in Scotland. This paper presents a techno-economic analysis quantifying the financial and environmental benefits of relocating latency-tolerant AI compute facilities from London to Kilmarnock, Scotland, to utilise the abundant wind resource. A comprehensive cost model was developed to contextualise energy related savings, covering computational hardware (Nvidia H100 GPUs), construction, land acquisition, and energy systems. Using NREL's REOpt software, various hybrid energy configurations—integrating grid, wind, solar, and battery storage—were optimised for a 51 MW facility. The results demonstrate that a wind-supported facility in Kilmarnock achieves a 16% reduction in total six-year development costs (£266 million savings) compared to a London-based grid-only baseline. Over a twenty-five-year horizon, the energy system savings alone amount to £647 million with a 1077kt reduction in carbon emissions. The study concludes that co-locating AI data centres with renewable generation offers a viable solution to grid curtailment while improving the economic feasibility of hyper-scale developments.

Keywords: *Techno-economic Analysis, Data Centres, Hybrid Energy Systems, Wind Curtailment, Artificial Intelligence.*

Symbol	Unit	Description
Financial & Taxation		
NPV	£ / \$	Net Present Value of costs or tax relief
PV_{6yr}	£ / \$	Present Value of an asset after a 6-year horizon.
T_{Corp} / T_c	%	Statutory Corporate Tax rate
r	%	Annual discount rate (cost of capital)
e	%	Annual escalation rate for O&M and Energy costs.
L	Years	Useful life of the asset or analysis period
n, t	-	Time indices for years
$RemCapEx$	£ / \$	Residual capital balance for writing-down allowances
Facility & Hardware		
$CapEx$	£ / \$	Initial Capital Expenditure
$OpEx$	£ / \$	Annual Operating Expenditure
P_{total}	MW	Total facility power requirement (Load)
P_{IT}	MW	Power consumed specifically by IT equipment
PUE	-	Power Usage Effectiveness (P_{total}/P_{IT})
N_{chips}	-	Total number of GPU units in the facility
C_{inst}	\$	Total installed cost per GPU (including networking)
C_{GPU}	\$	Market acquisition price of a single GPU chip
C_{DGX}	\$	Cost of a single DGX server unit
N_{DGX}	-	Number of GPUs per DGX server unit
Energy Systems		
C_{elec}	\$/yr	Total annual cost of electricity
$C_{grid\ avg}$	\$/MWh	Average price of grid-supplied electricity.
$LCOE$	\$/kWh	Levelised Cost of Energy for on-site generation
CP	%	Capacity Factor of renewable generation (Wind/Solar)

1 Introduction

In 2024 alone the UK spent £393 million to curtail 8.3TWhs of operators’ wind production, paying companies to cease energy generation to prevent grid overload [27]. Over 90% of these curtailment payments went to Scottish operators due to region’s large wind generation capacity and low population density. This paper aims to address this inefficiency by leveraging a different emerging challenge.

Roughly 4% of the United Kingdom’s power supply is currently used to power data-centres, 10-12 TWh annually [3]. This figure is projected to triple over the next 10 years to 33 TWh by 2035. This rapidly increasing demand could be alleviated by the aforementioned 8.3TWh of wind generation currently being curtailed at huge cost to the tax payer. At present, however, the majority of the UK’s data centres are located far away from this curtailed energy, with 42% of the country’s data centres located in London alone and only 5.6% in Scotland [7].

London’s high data centre density is not without reason, not all data centres serve the same purpose, many benefit from co-locating with the clients they serve to minimise latency; making London and other major English cities obvious locations. However, the high-growth data centre industry is increasingly driven by recent developments in Artificial Intelligence. These facilities comprise massive data centres with significantly larger power demands [18]. Crucially, these AI workloads benefit less from extremely low latency to their clients and can therefore be located more remotely. As grid connections become more competitive and metropolitan grids face extreme demand, remote locations have become more attractive to operators [17]. With OpenAI announcing a 230MW data-centre to run of hydropower in northern Norway [25], placing data centres next to the clean cheap energy they require is becoming an increasingly attractive strategy.

These developments have spurred energy research groups to build hybrid energy system models exploring energy savings and improved security. However, these papers often lack the specificity required for data centre development and fail to contextualize savings, resulting in a lack of reliable, commercially available pricing models for data centres and their accompanying energy systems.

This techno-economic analysis aims to fill that gap by developing a model for data-centre levelised development costs, ranging from GPU procurement to energy system installation and operation. The study aims to determine the potential financial savings of relocating AI compute data centres from London to Scotland, potentially lightening the burden of annual curtailment costs while spurring growth in a region that is too often overlooked. Specifically the paper will compare a data centre in London to one based south of Glasgow, in Kilmarnock; due to the regions large, pre-existing, wind farm capacity and fibre optic internet connection.

2 Methodology

2.1 Hardware

Hardware, in this context, refers to the components in the data-centre that provide the computing power necessary for the data-centre to carry out its function; Whether that be training AI models, hosting servers or running simulations.

2.1.1 Acquisition Cost

To evaluate the cost of hardware the method outlined in ‘The Rising Costs of Training Frontier AI Models’ [6] was followed.

Four GPUs were evaluated initially, the H100, GB200, H200, B200, all of which are Nvidia GPUs and were chosen for being some of the most common, commercially available, chips for current and planned hyperscale data centres. In particular, the H100 has been the most commonly used GPU for existing hyperscale data centres and is the GPU chosen moving forward [16]. The acquisition costs for these GPUs are difficult to estimate with lasting precision due to the volatility and commercial sensitivity of the prices. Therefore, this section should be viewed as outlining a methodology that must be regularly updated as prices change.

$$C_{\text{inst}} = \frac{C_{\text{GPU}} + \frac{N_{\text{DGX}} C_{\text{GPU}} - C_{\text{DGX}}}{N_{\text{DGX}}}}{0.81} \quad (1)$$

The acquisition costs are derived by compiling averages from multiple sources and adjusting for inflation. The Intra-server costs are derived by dividing the DGX price by the price of the GPUs contained in the DGX. The Inter-server cost is then assumed to be 19 percent of the total GPU acquisition cost, the total GPU installed cost can therefore be calculated using the equation below. When DGX prices could not be found a multiplier of 1.54x [6] the chip acquisition price was used for Nvidia GPUs [15, 1, 14, 20]. The results of the acquisition cost evaluation are shown in [3].

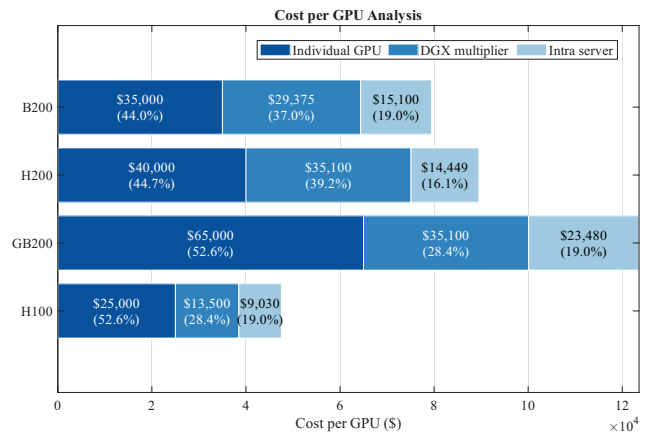


Figure 1: Cost per GPU Analysis

2.1.2 Energy Usage

To find the number of chips required, and therefore the cost of the hardware, a uniform power load of 51 MW was used. This value was chosen to align with the EpochAI GPU cluster database and to match similar Techno-Economic Analyses (TEAs) [28]. The total system power of an H100 DGX can be found on its data sheet to be 10.2 kW [24]; since there are 8 GPUs in an H100 DGX, this equates to 1.275 kW per GPU. To find the total installed power usage per chip, however, the power used in the data centre for non-computing tasks (cooling, lighting) must be accounted for. This is calculated using the Power Usage Effectiveness (PUE):

$$PUE = \frac{P_{\text{total}}}{P_{\text{IT}}} \quad (2)$$

For the example data centre, a PUE value of 1.1 was used, taken from Google’s trailing 12-month average [11]. With this, the total power requirement per chip and the number of chips in a 51 MW data centre can be calculated:

$$N_{\text{chips}} = \frac{P_{\text{total}}}{\left(\frac{P_{\text{DGX}}}{N_{\text{DGX}}}\right) \times PUE} \quad (3)$$

Using Equation 3 the 51 MW example data centre was found to require 36,364 H100 GPUs. With the installed acquisition cost per H100 determined to be \$47,530, making total hardware acquisition CapEx is \$1,728 million.

2.2 Building

In order to house, cool and protect the computational hardware a physical data centre building must be constructed. To calculate the price of the building, the Turner and Townsend [31] Data centre construction index was used. The London cost index was used for the building costs in London, and the Cardiff cost index used as a proxy for Kilmarnock, as no other British index was available. Due to the scale and rack density of the proposed data centre, it is assumed to be liquid-cooled; therefore, in line with the Turner Townsend index, an 8% increase is added to the air-cooled indices for both locations. A summary of the building costs can be found in Table 1.

Table 1: Building Costs Breakdown

Category	London	Rest of UK
Liquid Cooled Cost Index [\$/W]	12.9816	12.6792
<i>Capital Expenditure (CapEx)</i>		
GC/GRs and fees (preliminaries) [M\$]	66.206160	64.663920
Core / shell and architectural [M\$]	59.585544	58.197528
Mechanical (incl. equipment) [M\$]	218.480328	213.390936
Electrical (incl. equipment) [M\$]	317.789568	310.386816
Total CapEx [M\$]	662.061600	646.639200
<i>Operating Expenditure (OpEx)</i>		
Annual Building O&M [M\$]	7.65	7.65
Annual OpEx [M\$]	7.65	7.65

2.3 Land

Determining the land requirement for a hyper-scale data centre is challenging. Companies rarely specify the exact land area required for their data-centres and those that do often fail to distinguish between the total project area and the “white space” (the area dedicated solely to server rooms). While some sources specify white space, others report total building space—including offices, security, and utility areas. Neither metric effectively indicates the total plot size required.

To estimate the land requirement for the proposed facility, seven data centres were selected whose power requirements and total land usage were publicly available. These values were plotted to determine the correlation between power capacity and land requirement, as shown in Figure 2.

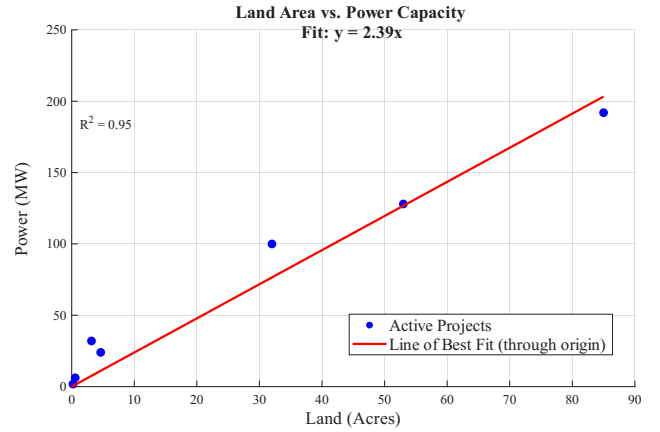


Figure 2: Land Area vs. Power Capacity Correlation

Using the correlation the land requirement for the example data-centre can be found using equation 4, the cost per acre for London is taken from the governments industrial land price appraisals 2019 using the slough price [19], and the Kilmarnock land price per acre is for pasture land in southern Scotland from Strutt and Parker [29]. An ideal comparison would be between industrial land in both locations, however, industrial land prices were difficult to ascertain in the Kilmarnock area due to its rural nature.

$$\text{Land}_{\text{Acres}} = \frac{P_{\text{total}}}{2.39} \quad (4)$$

Table 2: Land Aquisition Costs

Category	London	Kilmarnock
Land Requirement [Acres]	21.3	21.3
Land Cost [\$/Acre]	2,955,000	4,597
Total Land Acquisition Cost [M\$]	63.1	0.0980

For quantifying energy asset land usage; a conservative power density of 20 MW per acre was adopted for the wind farm layout to ensure ample spacing between turbines. Given this generous allocation, the

significantly smaller footprints required for the solar array and battery storage system were assumed to be co-located within the existing wind farm boundaries.

2.4 Energy

To optimise and price the hybrid energy systems powering the example data centre, the National Renewable Energy Laboratory’s (NREL) REOpt software was used [22]. REOpt is an optimisation tool built on Julia, which uses meteorological and economic data to model optimal energy production systems for specified asset energy requirements. Four energy system configurations were considered. All four systems assume a uniform load of 51 MW across the entire analysis period. It is assumed that 51 MW must be supplied at all times, with no permissible power outages. The grid is assumed to be 100% reliable for the duration of the model, meaning no grid outages.

2.4.1 Grid

Calculating the annual cost of electricity for the London base case is straightforward, as 100% of the electricity is drawn uniformly from the Grid. Therefore the annual cost can be found using the below equation, the energy price was taken from the governments average industrial energy price for a ‘Very Large’ asset [9]. The annual cost is calculated using Equation 5:

$$C_{\text{elec}} \left[\frac{\$}{\text{Year}} \right] = C_{\text{grid avg}} \left[\frac{\$}{\text{MWh}} \right] \times P_{\text{total}} [\text{MW}] \times 8760 \quad (5)$$

For the Kilmarnock cases, where on-site production assets are involved, the grid energy requirement varies dynamically. Fundamentally, as wind or solar generation increases, grid demand decreases. Therefore, a dynamic hourly energy price is required. The data used for this is from Octopus Energy’s Agile Tariff [10], which tracks the wholesale price, in this case in southern Scotland. The first 200 hours of this tariff is plotted in Figure 3.

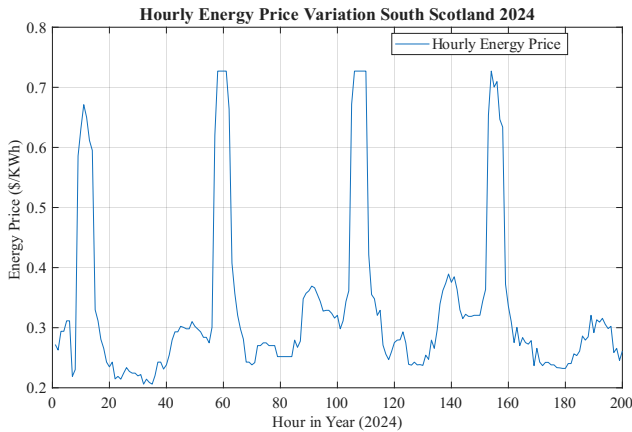


Figure 3: Kilmarnock Hourly Energy Price

Significant price fluctuations occur throughout the day, often in predictable patterns; the REOpt software

leverages these patterns to minimize costs. This is discussed in greater detail in the Results section.

2.4.2 Wind

The capacity factor data for wind production in Kilmarnock was obtained from the Renewables Ninja platform [26]. Table 3 specifies the inputs used to retrieve the capacity factor data plotted in Figure 4.

Table 3: Renewables Ninja Simulation Parameters

Parameter	Value
Dataset	MERRA-2 (global) (2024)
Capacity	1
Hub Height	80
Turbine Model	Vestas V90 2000
Latitude	55.6243
Longitude	-4.4595

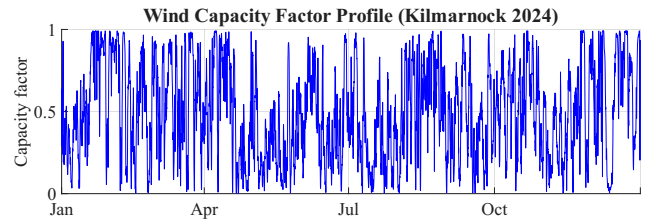


Figure 4: Kilmarnock Hourly Capacity Factor

Economic data for pricing the wind turbines was sourced from the UK government’s BEIS Electricity Generation Costs 2020 report [8]. The Wind costing data submitted to the REOpt solver are summarised in Table 5. As the system is optimised for a UK context, the MACRS schedule and US-specific federal incentives within REOpt were disabled; UK-specific tax benefits were applied to the NPV calculations manually outside of the solver, in line with Section 2.5.

Table 4: System Economic and Wind Simulation Parameters

Parameter	Value
Size Class	Large
System Capital Cost [\$/kW]	1561.26
O&M Cost (\$/kW per year)	38.35
Incentives	None
MARCS Schedule	None

2.4.3 Battery

The battery storage system consists of lithium ion battery packs connected to the grid, data-centre, and the on-site energy production assets. Table 5 presents the

battery pricing data used to inform REOpt. The Default REOpt efficiency values were used to calculate the battery energy losses [5, 21, 23].

Table 5: Battery Pricing Data

Parameter	Value
Energy Capacity Cost [\$/kWh]	115
Power Capacity Cost [\$/kW]	532
Constant Cost	222,115
Incentives	None
MACRS Schedule	None

2.4.4 Solar

Solar capacity factors were also derived from Renewables Ninja using the same coordinates. Table 6 below specifies the software inputs and the resultant capacity factor data is plotted in Figure 5.

Table 6: System Economic and Solar Simulation Parameters

Parameter	Value
Dataset	MERRA-2 (global) 2024
capacity	1
Combined loss fraction	0.1
Tracking	None
Tilt [°]	35
Azimuth [°]	180
Latitude	55.6243
Longitude	-4.4595
Mean capacity factor	11.20%

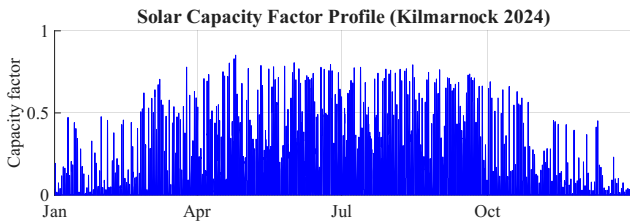


Figure 5: Kilmarlock Hourly Solar Capacity Factor

The solar farm pricing data, summarized in Table 7, was sourced from the same BEIS report used for wind assets.

Table 7: Solar PV Economic Parameters

Parameter	Value
Size Class	Large
System Capital Cost [\$/kW]	698
O&M Cost [\$/kW per year]	13
Incentives	None
MACRS Schedule	None

2.5 Financial

All asset costs are calculated as a six-year Net Present Value (NPV) to align with the estimated useful life of the computational hardware [2]. Energy systems are additionally compared over a 25-year horizon to match the REOpt solver's analysis period and evaluate the systems over their full operational lives. The NPV of an asset is determined by applying relevant tax relief and discounting future cash flows to present value, accounting for inflation, opportunity costs, and project risk through a discount rate. Table 8 specifies all financial metrics used [13, 30, 2].

Table 8: Financial and Simulation Metrics

Parameter	Value
Analysis Period	6
Energy System Comparison Period	25
Discount Rate	8.30%
Host Effective Tax Rate	25%
O&M Cost Escalation Rate	2.50%

2.5.1 Tax

It is assumed that the developer maintains sufficient taxable profit such that tax relief is realised in the year the expense is incurred. It is assumed that each years operational expenses will be expensed at the end of each respective year. Operational expenditures (OpEx) are expensed at the end of each respective year. The NPV of OpEx is calculated as follows:

$$NPV_{OpEx} = \sum_{t=1}^n \frac{OpEx_1(1 - T_c)(1 + e)^{t-1}}{(1 + r)^t} \quad (6)$$

Each assets CapEx will be taxed according to its specific HMRC taxation category. Table 9 below summarises which taxation category each asset will fall under [12].

Table 9: Building Tax Categories

Asset	Value	Asset Useful Life
Solar	Special	25
Wind	Special	25
Battery	Special	25
Hardware	Main	6
(Electrical)	Main	25
(Mechanical)	Main	12
(Shell)	Buildings	30
(Preliminaries)	Main	N/A
Land	None	N/A

The tax relief for each category is calculated using the following equations for Main Rate, Special Rate, and Structures and Buildings Allowance (SBA), respectively:

$$NPV_{Main} = \frac{CapEx \cdot T_{Corp}}{1 + r} \quad (7)$$

$$NPV_{Special} = \frac{0.5 \cdot CapEx \cdot T_{Corp}}{1 + r} + \sum_{n=2}^L \frac{0.06 \cdot T_{Corp} \cdot (0.94 \cdot RemCapEx_{n-1})}{(1 + r)^n} \quad (8)$$

$$NPV_{Build} = \frac{T_c \cdot 0.03 \cdot T_{Corp}}{1 + r} + \sum_{n=2}^L \frac{0.03 \cdot T_c \cdot (0.97 \cdot RemCapEx_{n-1})}{(1 + r)^n} \quad (9)$$

2.5.2 Depreciation

To quantify the total cost over the analysis horizon, the residual value of the assets must be considered. For this study, the total six-year NPV of an asset is calculated by subtracting the present value of the asset at year six and the total tax relief from the initial cost:

$$PV_{6yr} = \frac{C_{Initial}}{(1 + r)^6} \quad (10)$$

$$NPV_{CapEx} = C_{Initial} - NPV_{TaxRelief} - PV_{6yr} \quad (11)$$

3 Results

3.1 Energy Production

The REopt software was used to model three distinct configurations. Each scenario generated an energy dispatch profile and identified the most cost-effective production mix as additional technologies were integrated.

3.1.1 Wind and Grid

The primary configuration allowed REopt to integrate wind turbines to support the data centre load. The solver optimised purely for Net Present Value (NPV) against a grid-only baseline, with no explicit carbon reduction incentives or provisions for selling excess generation back to the grid. Figure 6 illustrates the energy profile for the first 500 hours of the year. While the simulation encompasses a full 8760-hour annual cycle, this close-up view highlights the optimizer’s hourly dispatch logic.

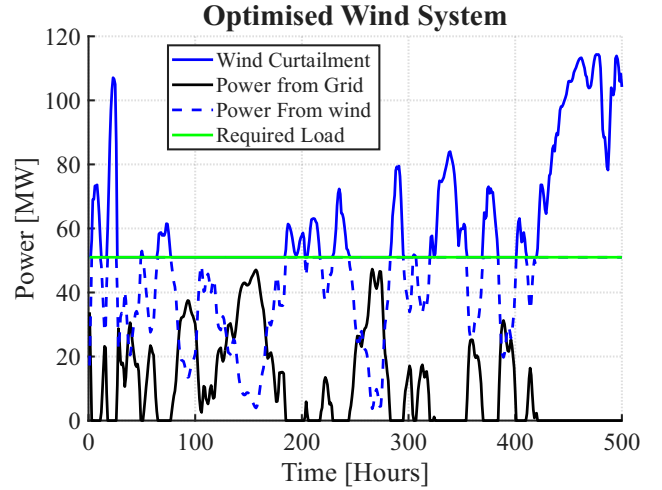


Figure 6: Wind and Grid 500 Hour Production Plot

The dispatch logic at this stage is clear: as wind generation increases, grid reliance decreases proportionally. REopt calculates the Levelised Cost of Energy (LCOE) for wind in this scenario at \$0.043/kWh—significantly lower than the Kilmarnock average grid price of \$0.21/kWh. The optimisation represents a trade-off between the costs of curtailed (unused) wind energy and the capital investment required for higher capacity to meet load during periods of low wind. Table 10 summarises the annual energy statistics.

Table 10: Wind and Grid Analysis Summary

Metric	Energy [GWh]	%
Total Energy Requirement	446.76	100.00
Wind Serving Load	348.72	78.05
Grid Serving Load	98.04	21.95
Wind Production Statistics	Energy [GWh]	
Total Wind Production	505.60	—
Wind Curtailment	156.89	—
Hybrid System Capacity	Power [MW]	
Wind Farm	115.3	

3.1.2 Wind, Battery and Grid

The addition of battery storage does not increase production, it adds flexibility allowing the system to store low-cost energy for dispatch during periods of high grid pricing. Figure 7 plots the production profiles for the first 100 hours of the year.

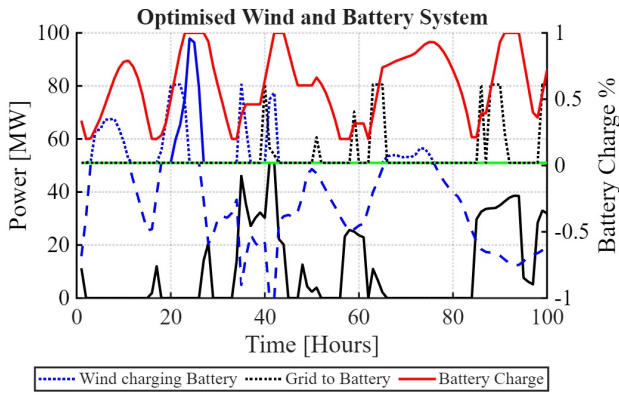


Figure 7: Kilmarnock Hourly Solar Capacity Factor

With the introduction of storage, the solver manages more complex variables. For instance, at approximately hour 20, surplus wind is directed to the battery until it reaches full state-of-charge, after which wind is curtailed. The solver also uses 'grid-to-battery' charging during low-price windows (e.g., hour 80), discharging the energy when grid prices peak. While the solver assumes perfect foresight of grid prices, these results demonstrate the potential for similar benefits using predictive modelling. Annual production data is summarized in Table 11.

Table 11: Wind, Grid and Battery Analysis Summary

Metric	Energy [GWh]	%
Total Energy Requirement	446.76	100.00
Wind Serving Load	322.81	74.49
Wind to Battery	18.78	4.20
Grid Serving Load	80.00	17.9
Grid to Battery	19.10	4.27
Round-Trip Battery Losses	-3.84	-0.86
Wind Production Statistics	Energy [GWh]	
Total Wind Production	445.41	–
Wind Curtailment	112.61	–
Hybrid System Capacity		
Wind Farm [MW]	105.934	
Battery [MW]	29.50	
Battery [MWh]	155.58	

3.1.3 Wind, Battery, Solar and Grid

The Capacity factors for solar are significantly higher during the summer months therefore the plot this time is taken from production data midway through the year. Figure 9 plots the dispatch data:

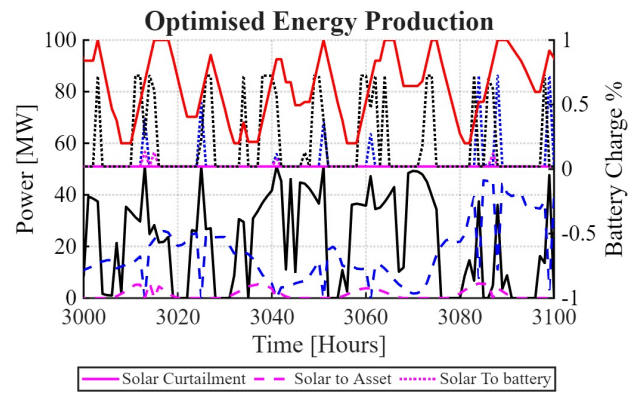


Figure 9: Kilmarnock Hourly Solar Capacity Factor

Solar generation predictably follows a daylight cycle, contributing to load and battery charging prior to evening grid price peaks. In winter months, this pattern remains consistent but with significantly reduced magnitude. Table 12 provides the full energy breakdown.

Table 12: Energy Analysis Summary

Metric	Energy [GWh]	%
Total Energy Requirement	446.76	100.00
Wind Serving Load	327.61	73.3
Wind to Battery	17.16	3.8
Grid Serving Load	73.82	16.5
Grid to Battery	18.69	4.2
PV Serving Load	12.05	2.69
PV to Battery	1.26	0.3
Round-Trip Battery Losses	-3.83	-0.86
Wind Production Statistics	Energy [GWh]	
Total Wind Production	454.61	–
Wind Curtailment	112.61	–
Solar Production Statistics	Energy [GWh]	
Total Solar Production	16.52	
Solar Curtailment	3.21	
Hybrid System Capacity		
Wind Farm [MW]	103.75	
Battery [MW]	28.80	
Battery [MWh]	151.90	
Solar Farm [MW]	16.83	

REopt calculates the Solar LCOE in Kilmarnock at \$0.08/kWh. While this is lower than grid prices, it is nearly double the wind LCOE. Consequently, while solar technology is economically viable, the meteorological conditions in Scotland make wind the significantly more attractive primary generation source. The results demonstrate that wind turbines provide the largest share of energy production, and therefore cost savings. While batteries provide critical flexibility and solar play a supplementary role for the Kilmarnock location.

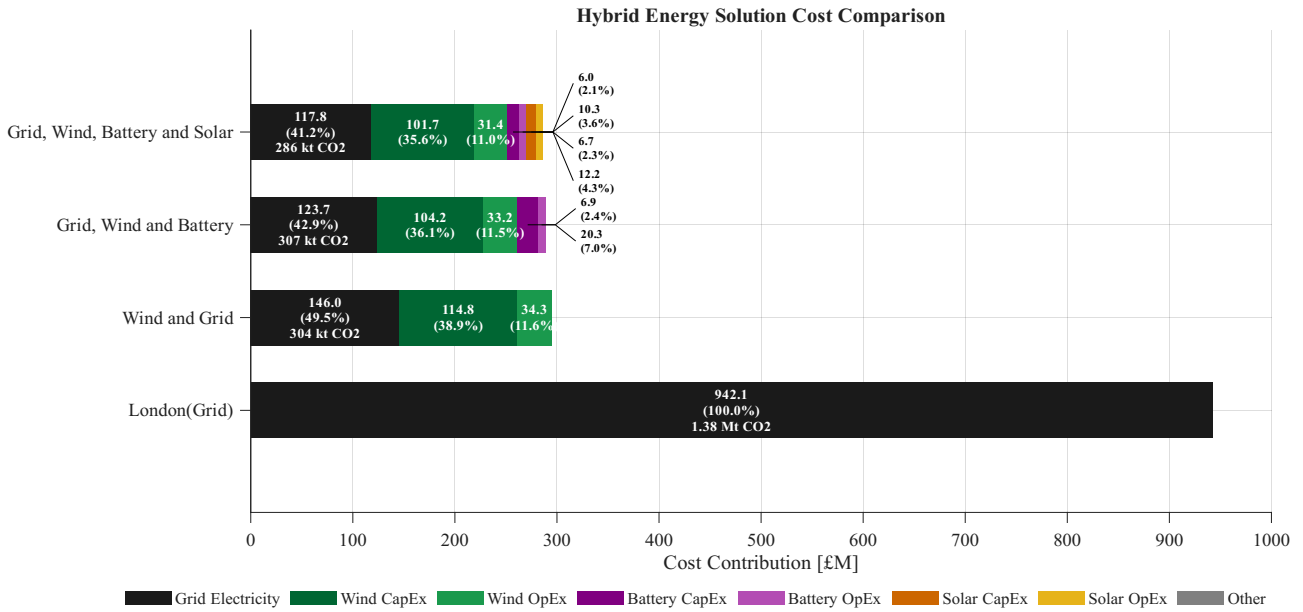


Figure 8: 25 Year Energy System Analysis

3.2 Economic Comparison

Using the pricing data and Net Present Value (NPV) calculations outlined in the methodology, each case was assessed over a 25-year period. The analysis first compared the energy systems in isolation, followed by an assessment alongside hardware, building, and land costs over a 6-year period to determine the total cost of ownership (TCO) for each scenario. Up to this point, all values within the RE-opt software were processed in USD; for these economic comparisons, they have been converted back to GBP at an exchange rate of 1.3135 [4].

3.2.1 25 Year Comparison

The 25-year NPV for each case was calculated assuming all energy assets fall within the special rate tax bracket. Figure 8 illustrates the economic benefits of integrating wind turbines into the generation mix, showing a saving of £647 million and a reduction of 1077 kt of CO_2 over the 25-year period.

Compared to the significant gains from wind integration, the addition of battery storage provided more modest but still notable savings of £7 million. Interestingly, the results showed a marginal increase in total carbon emissions for the battery-inclusive case due to higher grid consumption. However, this result may be slightly skewed: while grid-to-load consumption decreased, grid-to-battery charging increased. In reality, the battery typically charges during periods of low grid prices, which in the UK correlates with high renewable penetration and lower carbon intensity. As this model used an annual average carbon intensity for all grid withdrawals, it likely overestimated the battery system's emissions.

Finally, solar generation resulted in an even more modest £2 million saving over 25 years but achieved

the lowest overall grid reliance and carbon footprint. While the financial impact of solar is all but negligible for a project of this scale, its environmental contributions remain significant. Ultimately, the data suggests that the primary economic driver for the hybrid system is the well understood wind production potential in the Kilmarnock area.

3.2.2 6 Year Full Data-centre Comparison

To provide a fair comparison between development locations, the energy economics must be contextualized within the broader scope of data centre CAPEX. Figure 10 takes the energy production economic data over a six year period along with the hardware, building and land costs to compare the full development costs in Kilmarnock and London.

While absolute savings between energy cases are naturally lower over six years than twenty-five, the Kilmarnock location offers significant cost advantages in land and building infrastructure. The total development cost for the Kilmarnock 'Wind' case is £266 million (16%) lower than the London baseline. Battery and solar additions contribute minor further reductions of £7 million (0.4%) and £0.8 million (0.04%) respectively. Notably, land costs appear to decrease as solar and battery assets are added; this is due to the conservative wind turbine spacing assumed in the methodology, which allows for these assets to be co-located within the existing turbine footprint.

4 Conclusion

The 25-year and 6-year comparisons consistently demonstrate that wind turbine integration at the Kilmarnock site produces substantial savings in both cost and carbon emissions. Even when viewed through the

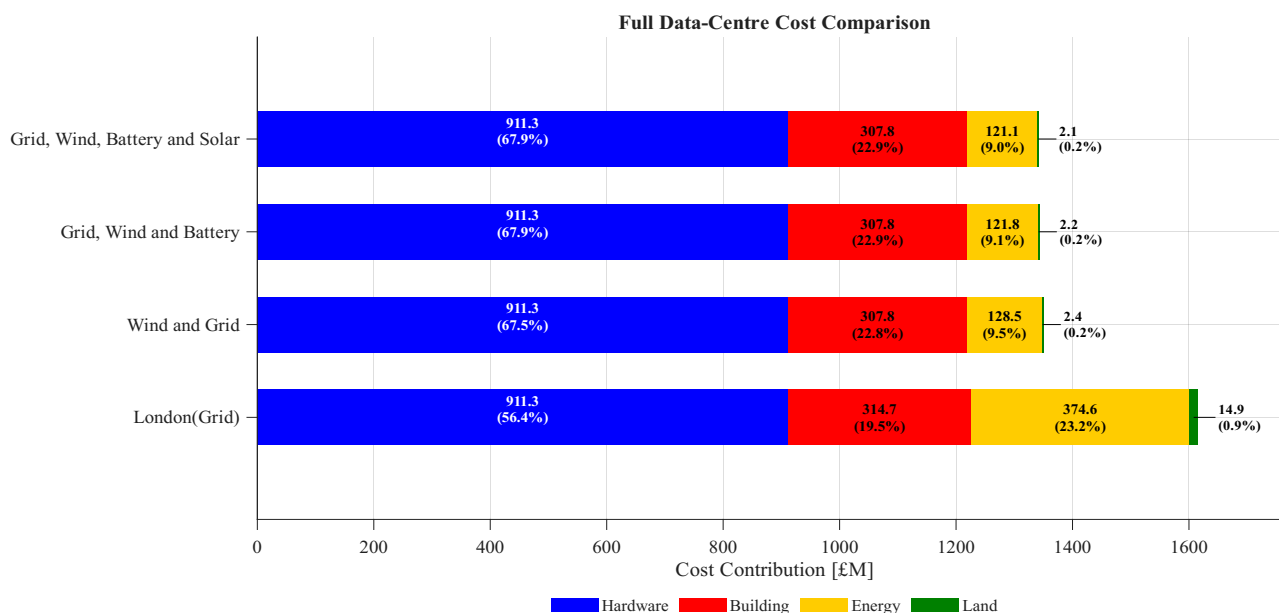


Figure 10: 6 Year data centre development costs

shorter 6-year window, these savings represent a significant portion of total development costs, before accounting for the substantial environmental benefits.

While the benefits of solar and battery additions are less pronounced for this specific location, this speaks to the utility of the model, identifying the real resource of the antiknock location to be wind. By updating the pricing and meteorological inputs, the solver can be used to assess any global location with sufficient accuracy to inform commercial or policy decisions.

Future iterations of this model could explore several promising avenues:

Grid Export: Allowing curtailed wind and solar generation to be sold back to the grid to further improve ROI.

Waste Heat Recovery: Utilizing low-grade heat from GPUs for district heating or industrial use, transforming a cooling cost into a potential revenue stream.

Price Prediction: Integrating a grid price prediction model to replace the current 'perfect foresight' assumption, providing a more realistic assessment of battery arbitrage potential.

The main objective of this model was to assess a location's suitability for hyper-scale data centre development, placing a hyper scale data centre in Kilmaronock shows clear quantified environmental and cost benefits even with all costs considered over a 6-year period. Even before accounting for the curtailment crisis ongoing in the UK, if direct policies are put in place to incentivise intelligent data centre development, in Scotland, the savings for the operators and the government will be even larger. Policy makers must move to take advantage of this to save the taxpayer money, contribute growth to a high priority industry and protect the environment.

References

- [1] AleksandarK. NVIDIA Blackwell GB200 Superchip to Cost up to 70,000 US Dollars, 5 2024.
- [2] Amazon.com Inc. Form 10-K: Annual Report for the fiscal year ended December 31, 2023, 2 2024.
- [3] Aurora Energy Research. Impact of Datacentres on the GB Power System. Technical report, Aurora Energy Research, 6 2025.
- [4] Bank of England. Daily Spot Exchange Rates against Sterling, 11 2025.
- [5] BloombergNEF. Lithium-Ion Battery Pack Prices See Largest Drop Since 2017, Falling to \$115 per Kilowatt-Hour, 12 2024.
- [6] Ben Cottier, Robi Rahman, Loredana Fattorini, Nestor Maslej, Tamay Besiroglu, and David Owen. The rising costs of training frontier AI models. 2 2025.
- [7] Data Center Map. Data Centers - Database, 2025.

- [8] Department for Business Energy & Industrial Strategy. Electricity Generation Costs 2020. Technical report, Government of the United Kingdom, 8 2020.
- [9] Department for Energy Security and Net Zero. Table 3.4.1: Prices of fuels purchased by non-domestic consumers in the United Kingdom excluding/including CCL. Quarterly Energy Prices (QEP), 2025.
- [10] Energy Stats UK. Octopus Agile Southern England 2025, 2025.
- [11] Google. Power usage effectiveness – Google Data Centers, 2025.
- [12] HM Revenue & Customs. Claim capital allowances: Overview, 2025.
- [13] HM Revenue & Customs. Corporation Tax rates and allowances, 2025.
- [14] Jarvislabs.ai. H200 Price Guide: Nvidia H200 GPU Cost & Rental Options (November 2025), 11 2025.
- [15] Jarvislabs.ai. NVIDIA H100 Price Guide 2025: Detailed Costs, Comparisons & Expert Insights, 10 2025.
- [16] James Sanders Lennart Heim Konstantin Pilz Robi Rahman. Data on GPU Clusters, 12 2025.
- [17] Tamsyn Lonsdale-Smith and Malachi Moses-Gair. Data Centre Impact Study. Technical report, National Grid DSO and Regen, 11 2025.
- [18] McKinsey & Company. AI power: Expanding data center capacity to meet growing demand. *McKinsey & Company*, 10 2024.
- [19] Ministry of Housing Communities & Local Government. Land value estimates for policy appraisal 2019. Technical report, Government of the United Kingdom, 5 2020.
- [20] Modal. How much does it cost to run NVIDIA B200 GPUs in 2025?, 7 2025.
- [21] Modo Energy. November 2024: GB battery energy storage research roundup, 11 2024.
- [22] National Renewable Energy Laboratory. REopt: Renewable Energy Integration & Optimization Web Tool, 2024.
- [23] National Renewable Energy Laboratory. Utility-Scale Battery Storage — Electricity — 2024 — ATB, 2024.
- [24] NVIDIA. Introduction to NVIDIA DGX H100/H200 Systems — NVIDIA DGX H100/H200 User Guide, 9 2025.
- [25] OpenAI. Introducing Stargate Norway, 7 2025.
- [26] Stefan Pfenninger and Iain Staffell. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. *Energy*, 114:1251–1265, 2016.
- [27] Renewable Energy Foundation. Discarded wind energy increases by 91% in 2024, 1 2025.
- [28] William Rollinson, Andrew Urquhart, and Murray Thomson. Technoeconomic Feasibility of Wind and Solar Generation for Off-Grid Hyperscale Data Centres. *Energies*, 18(2), 1 2025.
- [29] Strutt & Parker. Scottish Farmland Market Review 2024, 2024.
- [30] The Renewables Infrastructure Group. 2024 Annual Results Presentation, 2 2025.
- [31] Turner & Townsend. Data centre cost trends - Data centre construction cost index 2025, 2025.

Global Sensitivity Analysis of Economic Parameters in CAR T-Cell Therapy Manufacturing

Marta García Belza and Rebecca Fayad

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Chimeric Antigen Receptor (CAR) T-cell therapies have demonstrated significant clinical efficacy in the treatment of haematological malignancies; however, their widespread adoption remains constrained by high costs. These challenges are further amplified by patient-specific variability inherent to autologous production and by limited understanding of how manufacturing uncertainties propagate to system-level performance. This study aims to develop representative reduced-order models of critical cost metrics for their integration as constraints within an overall supply-chain optimisation framework. A variance-based Global Sensitivity Analysis (GSA) was performed on an extensive set of cost parameters which drive three Key Performance Indicators (KPIs). Factor Prioritisation (FP) was subsequently conducted to identify critical parameters for each KPI, allowing a significant reduction in problem dimensionality. Finally, a linear regression surrogate model was constructed using the selected parameters, which proved to accurately capture and predict the economic variability observed in the manufacturing process simulation. This work demonstrates that reduced-order cost models derived from detailed manufacturing simulations can effectively assess production feasibility while substantially reducing computational demand, enabling efficient and tractable optimisation for scalable, patient-centric CAR T-cell delivery.

Keywords: *CAR T-Cell Therapy Manufacturing – Economic Analysis – Global Sensitivity Analysis (GSA) – Factor Prioritisation – Linear Regression – Supply Chain Optimisation*

1 Introduction

Cancer immunotherapy has transformed the treatment landscape by harnessing the immune system to achieve targeted and durable anti-tumour responses^[1]. Among these approaches, Chimeric Antigen Receptor (CAR) T-cell therapy represents a major advancement through its personalised and genetically engineered treatment strategy. In this approach, patient-derived T-lymphocytes are genetically modified *ex vivo* to express recombinant CARs that redirect antigen recognition and cytotoxic activity towards malignant cells^[2]. This individualised strategy has demonstrated substantial clinical efficacy, particularly in B-cell malignancies, where CAR T-cell therapies have achieved high remission rates in patients with relapsed or refractory disease^{[3][4]}. Since the first regulatory approvals in 2017, these clinical successes have driven rapid growth in both academic research and commercial interest^[1]. However, the autologous nature of CAR T-cell therapy introduces significant manufacturing, logistical, and cost-related challenges, highlighting the need for systematic analysis and optimisation of production and delivery frameworks^[5].

2 Background

Chimeric Antigen Receptor (CAR) T-cell therapies have progressed from early clinical studies to approved treatments for several haematological

malignancies. Despite strong clinical outcomes, widespread access remains limited by the complexity and cost of autologous manufacturing, in which each therapy is produced as a patient-specific batch under strict time and quality constraints. This manufacturing paradigm tightly couples clinical scheduling with production capacity and requires stringent control of chain of identity, chain of custody, and regulatory compliance^{[5][6]}.

CAR T-cell therapy manufacturing is performed in batch mode through sequential unit operations including cell collection, processing, gene transfer, expansion, formulation, and release testing^[7]. The autologous nature of production restricts volumetric scale-up and economies of scale, requiring demand to be met through scale-out via parallel processing^{[2][5][6]}. Consequently, manufacturing costs increase approximately linearly with patient throughput, making economic performance highly sensitive to process efficiency and resource utilisation rather than production volume^[5].

Manufacturing represents the dominant cost contributor for CAR T-cell therapies, driven by specialised raw materials, single-use consumables, labour-intensive operations, cleanroom infrastructure, and extensive quality assurance requirements^{[2][5]}. Prior studies have shown that variable operating costs are largely driven by high-value reagents, while fixed and capital costs are dominated by facility design and parallel processing capacity^{[2][8]}. Long process durations

and bottleneck operations further reduce facility utilisation, increasing cost per therapy.

While Triantafyllou et al.^[6] and prior studies have characterised CAR T-cell therapy manufacturing systems and proposed optimisation frameworks under deterministic assumptions^{[2][5]}, current manufacturing cost analyses provide limited insight into how variability in individual process parameters affects economic performance. Health economic studies have instead focused on cost-effectiveness and pricing from payer perspectives, offering limited resolution on internal manufacturing cost drivers^[9].

Recent advances in process modelling have enabled detailed simulation-based estimation of CAR T-cell therapy manufacturing costs using platforms such as SuperPro Designer^[7]. However, the computational expense of these models limits their direct application in optimisation. This motivates this work, which identifies key economic drivers within manufacturing using variance-based Global Sensitivity Analysis (GSA) and reduced-order surrogate modelling to enable efficient economic analysis and optimisation.

3 Methodology

To address the identified gap, the following methodology was developed and applied:

- i) Identification of input parameters influencing each economic performance metric (KPI).
- ii) Reduction of problem dimensionality through the selection of critical parameters using variance-based GSA.
- iii) Modelling of dependencies via the construction of reduced-order surrogate models to capture parameter-cost relationships and quantify economic uncertainty.

3.1 SUPERPRO DESIGNER SOFTWARE

An existing SuperPro Designer® model of autologous CAR T-cell therapy manufacturing was used in this study. The model employs the built-in flowsheet and represents the full manufacturing workflow. SuperPro Designer is a steady-state process simulation software in which each flowsheet evaluation corresponds to a fixed set of input parameters and produces deterministic mass, energy and economic outputs. A single simulation typically requires 2-20 seconds, enabling efficient generation of large input-output datasets for sensitivity analysis^[7]. As part of this study, the SuperPro Designer software was connected to Python via the Component Object model (COM) Interface^[10], for the automatic iteration of varying parameter configurations, eliminating the need for manual user input for each case.

3.2 PARAMETER IDENTIFICATION

3.2.1 Key Performance Indicators (KPIs)

The economic analysis focuses on three key performance indicators (KPIs): capital expenditure (CAPEX), operating expenditure (OPEX), and Unit Cost per therapy. These metrics are standard indicators used to evaluate economic feasibility and performance in biopharmaceutical and advanced therapy manufacturing, including autologous cell and gene therapies^{[2][6][8][11][12]}. Together, they capture fixed investment requirements, recurring operational costs, and the per-therapy economic burden associated with patient-specific production.

Economic KPIs are computed for each flowsheet simulation using the built-in cost evaluation interface within SuperPro Designer®, which implements standard process economics accounting methods^[10]. Capital expenditure is calculated as the total capital investment (TCI), defined as the sum of direct fixed capital and working capital:

$$\text{CAPEX} = \text{TCI} = \text{DFC} + \text{WC} \quad (1)$$

where direct fixed capital (DFC) includes costs associated with equipment, installation, instrumentation, and facilities, and working capital (WC) accounts for funds required to support ongoing operations^{[7][13]}. Start-up, validation, R&D costs, and royalties are excluded.

Operating expenditure is computed as the total annual operating cost, incorporating all recurring expenses associated with manufacturing operations:

$$\text{OPEX} = C_{\text{mat}} + C_{\text{cons}} + C_{\text{lab}} + C_{\text{util}} + C_{\text{waste}} + C_{\text{QC/QA}} + C_{\text{trans}} \quad (2)$$

where the individual terms represent costs associated with raw materials, consumables, labour-dependent activities, utilities, waste treatment and disposal, quality control and quality assurance, and transportation, respectively^{[2][7][8]}.

The Unit Cost per therapy is then obtained by annualising capital and operating costs and normalising by the annual production throughput:

$$\text{Unit Cost} = \frac{\text{CAPEX}_{\text{annualised}} + \text{OPEX}}{N_{\text{therapy}}} \quad (3)$$

where N_{therapy} denotes the number of therapies produced per year. This formulation reflects the inherently patient-specific nature of autologous CAR T-cell therapy manufacturing, where per-therapy cost is strongly dependent on process efficiency and facility utilisation^{[5][12]}.

Parameter Type	Parameter	Nominal Value (\$/unit, item, kg)	Lower Bound (\$/unit, item, kg)	Upper Bound (\$/unit, item, kg)
Equipment	FT 101 Thawing	10000	5000	15000
	CPS 101 Cell Washing 1	30000	15000	45000
	GPB 101 Activation and Transduction	19500	9750	29250
	RM 101 Cell Expansion	95000	47500	142500
	MF 101 Microfiltration	27040	13520	40560
	CPS 201 Cell Washing 2	50000	25000	75000
	DFT 201 Freezing	80000	40000	120000
Consumables	Cell Processing Disposable Kit	500	250	750
	2 L Cell Bag	500	250	750
	20 mL Cryobag	28	14	42
	250 mL Gas Permeable Bag	72	36	108
Materials	Viral Stock	498098.77	240000	750000
	Leukopak	7597.20	3700	11400
	Antibody Complex Mix	497869.79	248000	747000
	Media	498.10	240	750
	PBS EDTA Buffer	489.01	240	740

Table 1: Cost variables investigated and uncertainty bounds applied in the Global Sensitivity Analysis, based on 2022 standard nominal values extracted from the standard SuperPro Designer process model.

3.2.2 Key Input Variables

Built-in cost analysis and project economic evaluation calculations from SuperPro Designer were used to identify model parameters contributing most strongly to variance in the three selected KPIs. Parameter screening was based on information extracted from the Economic Evaluation Report and the Itemized Cost Report, which provide capital and operating cost breakdowns at both process, section and unit-operation levels.

This explicit cost breakdown allowed for the identification of variables with highest percentage contributions to total capital and operating costs, used in the initial stages of the economic study. Facility-dependent costs, raw materials, and consumables were identified as the dominant contributors to overall economic performance, while labour-dependent and laboratory costs were excluded due to limited user-level controllability and model uncertainty. Consequently, this study focuses on parameters associated with equipment units, stock materials, and consumables within selected unit operations. A maximum of 16 parameters were selected as a balance between computational tractability and sufficient representation of dominant cost drivers.

All parameters were defined relative to a standard industrial flowsheet configuration, using nominal values referenced to 2022 data for all user-specified inputs, including purchasing prices (Table 1). Each flowsheet evaluation represents the manufacture of a single batch, corresponding to one personalised therapy, with an assumed annual throughput of 962 batches.

Parameter uncertainty was represented by symmetric bounds of $\pm 50\%$ around nominal values extracted from the SuperPro Designer flowsheet. This range

is consistent with established practice in variance-based sensitivity analysis when detailed statistical distributions or empirical uncertainty data are unavailable, while remaining sufficiently broad to capture plausible variability in process and cost parameters^{[14][15]} potentially arising from supplier pricing, process performance, and operational assumptions such as equipment sizing.

3.3 REDUCING PROBLEM DIMENSIONALITY VIA GLOBAL SENSITIVITY ANALYSIS

Global sensitivity analysis (GSA) evaluates the influence of uncertain input parameters over their full feasible ranges, rather than relying on local perturbations around a nominal operating point^[14]. This enables a comprehensive assessment of how variability in manufacturing cost parameters propagates to economic outputs. By construction, variance-based GSA is applicable irrespective of model linearity or monotonicity, providing a robust and model-independent basis for sensitivity assessment and uncertainty propagation^{[16][17][18]}.

3.3.1 Variance Decomposition & Sobol Sensitivity

For a model output $Y = f(\mathbf{X})$ with $\mathbf{X} = (X_1, \dots, X_d)$, the total variance can be decomposed into contributions from individual inputs and their interactions^[16]:

$$\text{Var}(Y) = \sum_{i=1}^d V_i + \sum_{i < j} V_{ij} + \dots + V_{1\dots d} \quad (4)$$

where the first-order (main-effect) contribution of input X_i is given by $V_i = \text{Var}(\mathbb{E}[Y | X_i])$, while higher-order terms such as V_{ij} capture interaction effects between two or more parameters. Normalised Sobol

sensitivity indices are defined as the ratio of each variance component to the total variance, enabling direct comparison of parameter importance across the input space^{[14][17]}. This variance decomposition framework makes Sobol analysis particularly suitable for techno-economic models, where input uncertainties span wide ranges and the functional relationships between inputs and outputs are not known *a priori*.

The accuracy of variance-based GSA depends on the convergence of Monte Carlo estimators for the variance terms. Under standard assumptions of finite output variance and independent input sampling, Sobol index estimators are unbiased and converge asymptotically as the sample size increases^[16]. In this study, Sobol quasi-random sequences were employed to achieve low-discrepancy and uniform coverage of the input domain as seen in Figure 1, which improves estimator efficiency relative to purely random Monte Carlo sampling.

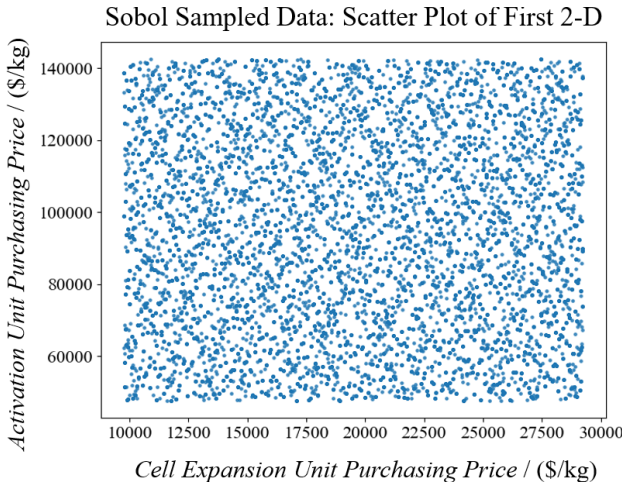


Figure 1: Sample scatter plot of two parameters, demonstrating uniformity of the input space using quasi-random Sobol sequences.

The sampling strategy proposed by Saltelli was implemented to efficiently compute variance-based Sobol sensitivity indices^[16]. This method provides a structured Monte Carlo sampling scheme that enables simultaneous estimation of first-order and total-effect Sobol indices with reduced computational cost compared to naive Monte Carlo approaches^[16]. Two independent sample matrices, $\mathbf{A}$ and $\mathbf{B}$ are generated, each of dimension $N \times d$, where N is the base sample size and d is the number of input parameters.

From these matrices, a set of hybrid matrices $\mathbf{A}_{\mathbf{B}}^{(i)}$ is generated by replacing the i th column of $\mathbf{A}$ with the corresponding column from $\mathbf{B}$. Model evaluations are then performed for $\mathbf{A}$, $\mathbf{B}$, and each hybrid matrix, resulting in a total of $N(d + 2)$ model runs. Then, using variance estimators derived from the sampled outputs, the first-order Sobol index for parameter X_i

is estimated as

$$S_i = \frac{\text{Cov}(Y_{\mathbf{B}}, Y_{\mathbf{A}_{\mathbf{B}}^{(i)}})}{\text{Var}(Y)} \quad (5)$$

while the total-effect index is estimated as

$$S_{T_i} = 1 - \frac{\text{Cov}(Y_{\mathbf{A}}, Y_{\mathbf{A}_{\mathbf{B}}^{(i)}})}{\text{Var}(Y)} \quad (6)$$

where $Y_{\mathbf{A}}$, $Y_{\mathbf{B}}$, and $Y_{\mathbf{A}_{\mathbf{B}}^{(i)}}$ denote the model outputs corresponding to the respective sampling matrices^[16]. These estimators quantify, respectively, the isolated contribution of each input and its total contribution including all interaction effects.

In this work, the SALib Python library was used to automate and implement Saltelli sampling^[19]. As the underlying SuperPro Designer model is deterministic, repeated evaluations under identical inputs produce identical outputs, and observed output variance can therefore be attributed solely to input uncertainty rather than numerical noise.

3.3.2 Sobol Indices and Uncertainty Propagation

Convergence and reliability of the Sobol estimates is assessed by verifying that the sum of the total-effect indices across all input parameters satisfies

$$\sum_{i=1}^d S_{T_i} \approx 1 \quad (7)$$

which indicates that the sampled input space sufficiently captures the total output variance^{[14][16]}.

The resulting Sobol indices were then used for Factor Prioritisation (FP)^[18] whereby parameters are ranked according to their contribution to output variance, and thus dominant cost drivers for each economic metric can be identified. Parameters with negligible total-effect indices are deemed as weakly influential and were excluded from further analysis via Factor Fixing (FF)^{[14][18]}. Those resulting in large first-order indices (S_i) were interpreted to have a strong direct influence on the KPI, whereas discrepancies between total-effect (S_{T_i}) and first-order indices were used to assess the presence and magnitude of interaction effects. A conservative threshold of $S_{T_i} \geq 0.005$ was established to retain a sufficiently broad parameter space while ensuring that only variables with a non-negligible contribution to output variance were included in the reduced model.

3.4 REDUCED-ORDER SURROGATE MODEL

Reduced-order surrogate modelling is appropriate when repeated evaluation of a high-fidelity simulator is computationally expensive and when downstream tasks require rapid model evaluations, such as uncertainty propagation, optimisation, and scenario screening^{[20][21]}. Within this framework, surrogate model selection is governed by the validity of underlying structural assumptions rather than model complexity alone. Linear regression is first considered due to its computational efficiency and interpretability when input-output relationships are approximately linear over the explored parameter space, multicollinearity is limited, and residual behaviour is well-behaved^{[22][23]}.

When these assumptions are violated, alternative surrogate models may be employed, including polynomial response surfaces for moderate nonlinearity^[24], kernel-based methods such as support vector regression for smooth nonlinear responses^[25], and artificial neural networks for highly nonlinear systems with complex interactions^[23]. As demonstrated in the Results section, the reduced manufacturing cost model satisfies the assumptions required for linear regression, and this approach is therefore adopted for subsequent analysis.

3.4.1 Linear Regression

Linear regression approximates the relationship between a set of input variables and a scalar output by assuming an affine functional form. Given a dataset of n_d observations $\{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^{n_d}$, where $\mathbf{x}^{(i)} \in \mathbb{R}^{n_x}$ denotes the input vector and $y^{(i)} \in \mathbb{R}$ the corresponding output, the regression model is defined as

$$\hat{y} = \theta_0 + \sum_{j=1}^{n_x} \theta_j x_j, \quad (8)$$

where θ_0 is the intercept and θ_j are the regression coefficients. In matrix form, this can be written as

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\theta}, \quad (9)$$

where $\mathbf{X} \in \mathbb{R}^{n_d \times (n_x+1)}$ is the design matrix augmented with a column of ones and $\boldsymbol{\theta}$ is the vector of unknown parameters^[26].

The parameters are estimated by minimising the sum of squared residuals between predicted and observed outputs, leading to the least squares problem

$$\min_{\boldsymbol{\theta}} f_{\text{obj}}(\boldsymbol{\theta}) = \sum_{i=1}^{n_d} \left(\mathbf{x}^{(i)\top} \boldsymbol{\theta} - y^{(i)} \right)^2, \quad (10)$$

which penalises large prediction errors and yields estimates that minimise the mean squared error^{[23][26]}.

Provided that the columns of $\mathbf{X}$ are linearly independent, this optimisation admits a closed-form solution given by the normal equations

$$\boldsymbol{\theta}^* = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (11)$$

This solution uniquely minimises the least squares objective and can be computed efficiently for moderate problem sizes^{[22][26]}. In surrogate modelling, this analytical tractability enables rapid model construction and evaluation, making linear regression well suited for repeated predictions in sensitivity analysis and optimisation.

3.4.2 Data Validation and Cross-Checking

To evaluate the predictive validity of the reduced linear regression surrogate, a train-validation data split was employed to assess out-of-sample performance. The SuperPro-generated dataset was partitioned into 70% training data and 30% validation data, a standard practice for assessing generalisation and detecting overfitting in surrogate modelling when independent data are unavailable^{[23][27]}. The model was fitted on the training subset and validated by comparing predicted outputs against SuperPro results for the validation set using parity analysis and standard error-based metrics^{[20][22]}.

The validity of the surrogate model depends on the fidelity of the underlying SuperPro Designer simulations. SuperPro Designer is a widely used process modelling platform whose economic calculations are grounded in established chemical engineering cost estimation methods and internally consistent mass and energy balances^{[7][13]}. The platform has been successfully applied to analyse alternative cell and gene therapy manufacturing strategies, yielding realistic and decision-relevant capital and operating cost estimates under representative process assumptions^{[5][6]}. Although SuperPro simulations are deterministic and steady-state and do not capture stochastic operational disruptions, they are well suited for comparative techno-economic and sensitivity-driven analyses.

4 Results & Discussion

4.1 GLOBAL SENSITIVITY ANALYSIS

An extensive global sensitivity analysis was performed on the 16 parameters outlined in Table 1 to understand their direct and combined contributions, rank their influences on the three KPIs investigated and discard those variables deemed less significant for the following modelling steps.

4.1.1 Sobol Sensitivity Analysis & Factor Prioritisation

A visual representation of the Sobol sensitivity indices is shown in Figure 2. It is evident by the intense red colour on the heat maps that a particular parameter has a much greater influence on all three KPIs over the other 15 parameters considered: the variability of the Cell Expansion Unit's purchasing price. Given its high first order index (S_i) of 0.90 and total order index (S_T) of 0.92, this parameter accounts for most of the uncertainty in the CAPEX's variance, as well as the OPEX and unit production cost with similar index values as outlined in Table 2.

(a) CAPEX Sobol Index Values

Parameter	S_i (%)	S_T (%)
Cell Expansion Unit	90.44	92.85
Microfiltration Unit	5.91	5.91
Activation and Transduction Unit	1.28	1.28

(b) OPEX & Unit Cost Sobol Index Values

Parameter	S_i (%)	S_T (%)
Cell Expansion Unit	90.32	89.90
Microfiltration Unit	5.77	5.77
Activation and Transduction Unit	1.23	1.23
2 L Cell Bag	6.04	6.05
Cell Processing Disposable Kit	2.37	2.42

Table 2: Dominant cost drivers for CAPEX, OPEX and Unit Cost per therapy, showing Sobol index values above the selected threshold of $S_T \geq 0.5\%$.

A user-set threshold^[6] of $S_T \geq 0.005$ was set for the selection of critical parameters for each of the outputs. This conservative value was chosen to ensure the highest accuracy and reliability of the dependencies modelled in the later steps, given less than 5 parameters were already performing above this. Setting the threshold to 0.01 would have restricted the selected parameter set to the three equipment unit purchase costs across all KPIs. However, this study deliberately adopted a lower threshold to ensure in-

clusion of relevant consumable-related parameters in the OPEX and unit cost analyses, thereby avoiding the omission of meaningful cost contributors. Further reduction of the parameter set is therefore recommended for future work.

The parameter values omitted from Table 2 performed well below this threshold, approximating zero in both their first and total order indices. This indicated their relatively minimal effects on the CAPEX, OPEX or Unit Cost per therapy, and thus were discarded as non-influential. This led to the eventual reduction in the size of the problem through Factor Prioritisation (FP)^{[14][16]}. The parameters selected for further use to model the output's dependencies are as outlined in Table 2: three equipment unit purchasing prices for the CAPEX, and the addition of two consumables' purchasing prices for the modelling of the OPEX and Unit Cost per therapy.

Minimal parameter interactions were identified given the total order sobol indices was simply equal to the sum of the first order indices for all parameters for both the OPEX and Unit Cost per therapy (Eq. 12), or of both the first and second order indices, as seen for the CAPEX in Figure 2 (Eq. 13).

$$\sum_{i=1}^d S_i \approx \sum_{i=1}^d S_{T_i} \quad (12)$$

$$\sum_{i=1}^d (S_i + S_{ij}) \approx \sum_{i=1}^d S_{T_i} \quad (13)$$

Thus, higher-order indices were not explored given the output variability was attributed to the direct effect of a single parameter on the output and their additive contributions^[15].

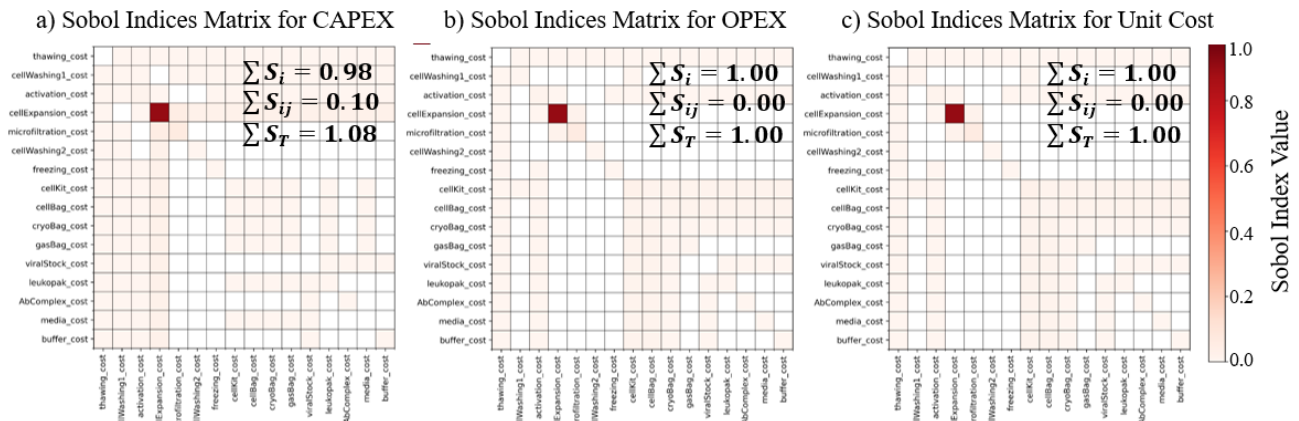


Figure 2: Representation of Sobol sensitivity indices via heat maps, illustrating the variance decomposition for each KPI to each specific variable, whether by direct effects (first-order S_i) or interactions between variables (second-order S_{ij}).

4.1.2 Reliability of Sensitivity Analysis

The reliability of the results was tested via three further approaches:

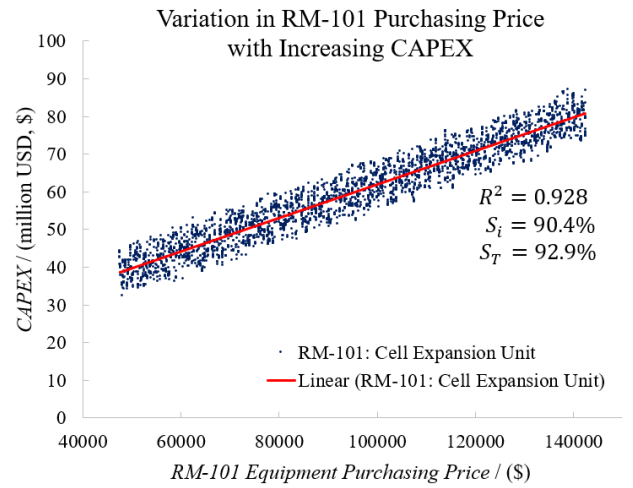
Convergence Test: The convergence of the variance-based sensitivity analysis was assessed by verifying that the sum of the total-effect Sobol indices satisfies $\sum_i S_{T_i} \approx 1$, as described in Section 3.3.2. An initial sampling effort of 4096 model evaluations was found to be insufficient, as the total variance was not fully captured. However, when performing a larger number of evaluations (8192), the condition was satisfied, proving that converged sensitivity estimates had been yielded^{[14][16]}. In accordance with standard Sobol sampling practice, the base sample size was selected as a power of two, $N = 2^k$, to improve quasi-random sampling efficiency and estimator convergence.

Visual Inspection: The trends identified through the sensitivity analysis were confirmed through visual inspection. Strong correlations were observed between the most influential parameters and the corresponding economic outputs as illustrated in Figure 3a, while parameters excluded from the reduced set exhibited negligible impact. Parameters with progressively lower contributions to the overall KPIs displayed increasingly wider output variability, reflecting weaker functional dependence. For variables with little to no influence, the resulting scatter plots consisted of randomly distributed points centred around a horizontal mean line (see Fig 3b), consistent with the expected behaviour under Sobol sampling. In addition, the best- and worse-case scenarios for each KPI was quantified (see Table 3) and seen to correspond to the parameter configurations with the lowest and highest bound values respectively, as expected. This provides a direct measure of the marginal impact of parameter variability on each economic metric.

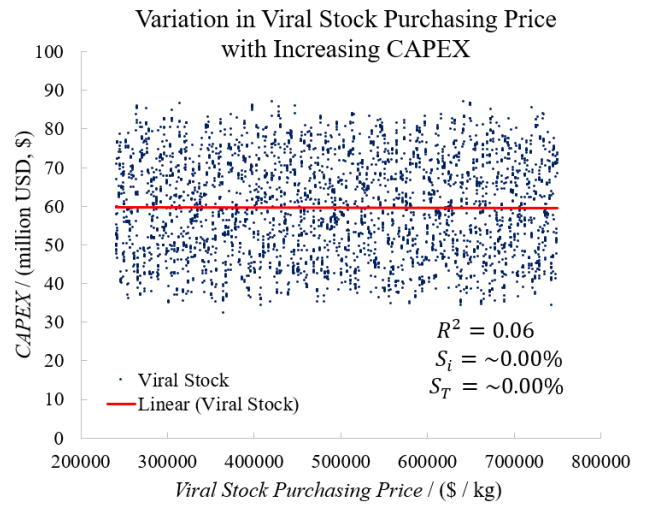
Process KPI	Nominal	Best-case	Worse-case
CAPEX (\$)	44,712,000	32,457,586	87,132,973
OPEX (\$/yr)	20,073,000	12,336,459	20,431,980
Unit Cost (\$/batch)	20,866	12,824	21,239

Table 3: Uncertainty in KPIs with the simultaneous variability in purchasing prices of +/-50%.

Independent Parameter Variation: To further isolate parameter effects, single-parameter variation analyses were performed. Each influential parameter was varied independently across its uncertainty range while holding all other inputs at their nominal, lower and upper values. The resulting percentage variability in CAPEX, OPEX, and Unit Cost corresponded with the trends observed, where Cell Expansion caused the CAPEX to double when comparing lower and upper bound values for this parameter, whereas stock material variabilities produced negligible impact on the KPIs.



(a) Most influential parameter.



(b) Non-influential parameter.

Figure 3: Covariance plots illustrating the impact on the CAPEX by (a) the most influential unit (RM-101) and (a) the least influential unit (Viral Vector), as identified via the Sobol sensitivity analysis, across all 16 parameters sampled input configurations.

4.2 REDUCED-ORDER SURROGATE MODEL

Based on the outcomes of the GSA, a linear regression surrogate model was selected to represent the economic behaviour of the manufacturing process, as the system was found to satisfy the key assumptions required for linear modelling. As previously stated, the Sobol sensitivity indices indicated that output variability is dominated by main effects, with limited higher-order interactions between input variables, supporting the use of an additive linear structure. Under these conditions, linear regression provides an interpretable and computationally efficient surrogate for capturing cost dependencies.

Following this justification, a reduced linear regression model was constructed for each economic output using the most influential manufacturing cost variables as outlined in Table 4.

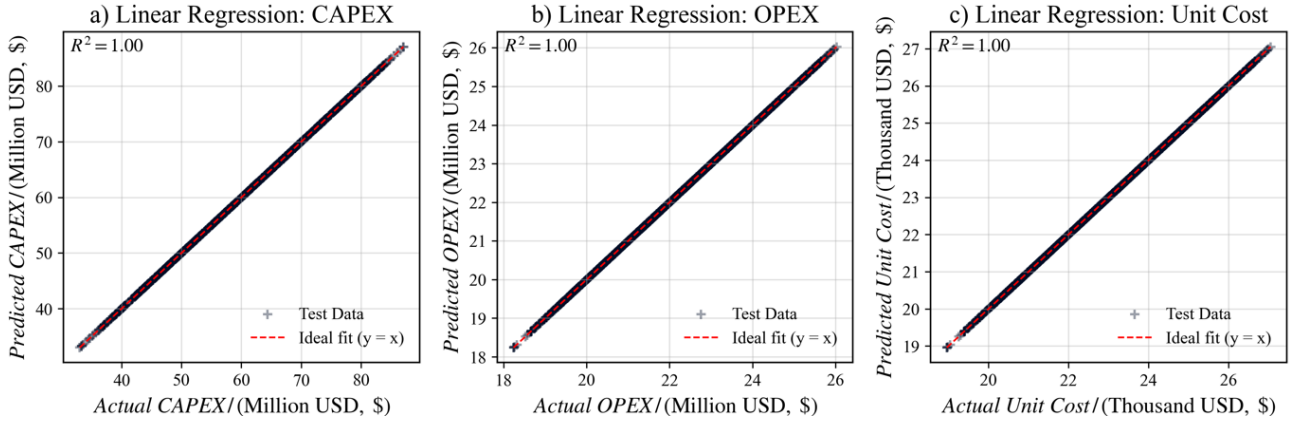


Figure 4: Plot of the linear regression formulated for the CAPEX, where a data ratio of 70:30 was used to train and test the model respectively. Uncertainty metrics are shown on the top-left corner, including the expression for the model and R-squared.

4.2.1 Linear Regression Model

The general linear surrogate model takes the form:

$$f_n(\mathbf{x}) = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_5 x_5 + z \quad (14)$$

where f_n denotes the economic output of interest, θ_i are regression coefficients, and z is the intercept term.

The fitted regression models for each KPI are:

$$f_1 = 253.0 x_1 + 442.7 x_2 + 392.4 x_3 + 2,658,357 \quad (15)$$

$$f_2 = 35.3 x_1 + 61.7 x_2 + 54.9 x_3 + 1924 x_4 + 962 x_5 + 12,769,300 \quad (16)$$

$$f_3 = 0.037 x_1 + 0.064 x_2 + 0.057 x_3 + 2.0 x_4 + 1.0 x_5 + 13,274 \quad (17)$$

In these expressions, f_1 , f_2 , and f_3 represent the predicted CAPEX, OPEX, and Unit Cost per therapy, respectively. All coefficients of variables seen in Table 4 quantify the marginal contribution of each input to the corresponding economic output within the defined uncertainty bounds.

Variable	Description	Unit / Equipment
x_1	Activation and transduction unit cost	GPB-101
x_2	Cell expansion unit cost	RM-101
x_3	Microfiltration unit cost	MF-101
x_4	Cell processing disposable kit cost	Consumable
x_5	Gas permeable bag cost	Consumable

Table 4: Definition of input variables used in the reduced surrogate model.

The regression results indicate that Cell Expansion unit's cost (x_2) is the dominant driver across all three economic metrics, exhibiting the largest coefficients across the KPI models. This finding is consistent with the central role of cell expansion in autologous CAR T-

cell therapy manufacturing, where extended process durations, specialised equipment requirements, and the need for parallel processing significantly influence both capital investment and operating costs.

Microfiltration (x_3) and Activation & Transduction unit costs (x_1) also contribute substantially to economic variability, reflecting the cost intensity and operational significance of upstream gene transfer and cell processing steps within the manufacturing process. In contrast, consumable-related variables (x_4 and x_5) appear only in the OPEX and Unit Cost models, with comparatively smaller coefficients, indicating that while consumables contribute to operating expenses, they do not affect capital investment.

4.2.2 Data Validation

The linear structure of the fitted models as seen in Figure 4 suggests that, within the investigated uncertainty bounds, the economic outputs respond approximately linearly to variations in the dominant cost parameters. The surrogate model was further validated using a hold-out cross-validation approach, previously outlined in Section 3.4.2. Parity plots comparing predicted and simulated outputs exhibited a one-to-one correspondence, with coefficients of determination ($R^2 = 1.00$) for all KPIs. This indicates that, within the sampled uncertainty space, the reduced linear model reproduces the SuperPro Designer outputs exactly, reflecting the appropriate selection of the reduced variables and suggesting that further dimensionality reduction may be feasible without loss of predictive accuracy.

These results support the use of linear regression as an effective surrogate modelling approach for this system, enabling rapid evaluation of economic performance and facilitating integration into downstream optimisation frameworks without repeated execution of the detailed process simulation.

5 Conclusion

Autologous CAR T-cell therapies have demonstrated transformative clinical outcomes for haematological malignancies, yet their widespread adoption remains constrained by high costs, complex manufacturing workflows, and tightly coupled vein-to-vein supply chains. The patient-specific nature of production limits economies of scale and amplifies sensitivity to process inefficiencies, resource utilisation, and cost variability across manufacturing operations. As a result, understanding how uncertainty in manufacturing parameters propagates into overall economic performance is critical for assessing feasibility and guiding scalable, patient-centric deployment of these therapies.

This study applied a structured GSA and reduced-order data-driven surrogate modelling approach to identify and quantify the dominant economic drivers in autologous CAR T-cell therapy manufacturing. By analysing uncertainty across 16 cost-related parameters, the work successfully ranked parameter influence on CAPEX, OPEX, and unit production cost, enabling a substantial reduction in model dimensionality without loss of predictive accuracy.

The analysis showed that economic variability is dominated by a small subset of equipment purchase costs, with the unit costs for cell expansion, micro-filtration, and activation & transduction emerging as the most influential contributors. Using these dominant variables, a reduced linear regression surrogate model was developed and validated against the SuperPro Designer process simulation, demonstrating consistent agreement while significantly reducing computational demand.

Overall, the results demonstrate that combining process simulation, sensitivity analysis, and surrogate modelling provides an efficient framework for identifying key cost drivers and enabling tractable economic analysis of CAR T-cell therapy manufacturing. This approach supports future optimisation efforts aimed at improving the economic feasibility and scalability of patient-specific cell therapy production.

6 Outlook

While this study focused on the economic drivers of the manufacturing process, future work should extend the integration between manufacturing models and existing vein-to-vein supply-chain optimisation frameworks, such as mixed-integer linear programming (MILP) models^[6]. Autologous CAR T-cell therapy production is inherently patient specific, and upstream biological variability, particularly differences in leukapheresis T-cell composition, can strongly influence unit operation performance, including acti-

vation efficiency, expansion kinetics, and overall processing times. Explicit incorporation of such patient-dependent variables into the manufacturing cost and time models would enable more realistic representation of supply-chain optimisation. This would allow for the explicit evaluation of trade-offs between cost, capacity, and clinically relevant turnaround times. Accounting jointly for biological variability and operational constraints is therefore essential to enhance the robustness and clinical relevance of patient-centric CAR T-cell supply-chain optimisation.

Moreover, the current analysis is based on deterministic steady-state simulations, so incorporating stochastic elements such as batch failure probabilities and reworks would make it more realistic and enable specific assessment of operational risk. The surrogate modelling framework could also be strengthened through further comparisons against alternative surrogate models under consistent and more elaborate validation criteria.

With this expanded set of cost and time inputs, surrogate model selection should be reassessed to account for increased nonlinearity and interaction effects. To facilitate integration within MILP frameworks, surrogate representations should remain optimisation-compatible, enabling efficient system-level optimisation without repeated evaluation of high-fidelity process simulations. As CAR T-cell therapies continue to demonstrate transformative clinical impact and expand across indications, such integratedscalable, and computationally efficient modelling frameworks will be essential to supporting economically sustainable delivery and broader patient access^[1].

References

- [1] Carl H. June et al. "CAR T cell immunotherapy for human cancer". In: *Science* 359.6382 (2018), pp. 1361–1365. DOI: 10.1126/science.aar6711.
- [2] Bruce L. Levine et al. "Global Manufacturing of CAR T Cell Therapy". In: *Molecular Therapy: Methods & Clinical Development* 4 (2016), pp. 92–101. DOI: 10.1016/j.omtm.2016.12.006.
- [3] Shannon L. Maude, Noelle Frey, Pamela A. Shaw, et al. "Chimeric antigen receptor T cells for sustained remissions in leukemia". In: *New England Journal of Medicine* 371.16 (2014), pp. 1507–1517. DOI: 10.1056/NEJMoa1407222.
- [4] Sattva S. Neelapu, Frederick L. Locke, Nancy L. Bartlett, et al. "Axicabtagene ciloleucel CAR T-cell therapy in refractory large B-cell lym-

- phoma". In: *New England Journal of Medicine* 377.26 (2017), pp. 2531–2544. DOI: 10.1056/NEJMoa1707447.
- [5] Maria M. Papathanasiou et al. "Autologous CAR T-cell therapies supply chain: challenges and opportunities?" In: *Cancer Gene Therapy* (2020). DOI: 10.1038/s41417-019-0157-z.
- [6] Niki Triantafyllou et al. "A digital platform for the design of patient-centric supply chains". In: *Scientific Reports* 12 (2022), p. 17365. DOI: 10.1038/s41598-022-21290-5.
- [7] Intelligen, Inc. *Cell & Gene Therapy Processes: Modeling and Evaluation with SuperPro Designer*. SuperPro Designer documentation and example report material. 2022.
- [8] Peter Vormittag et al. "A guide to manufacturing CAR T cell therapies". In: *Current Opinion in Biotechnology* 53 (2018), pp. 164–181. DOI: 10.1016/j.copbio.2018.01.025.
- [9] Roland Heine et al. "Health Economic Aspects of Chimeric Antigen Receptor T-Cell Therapies for Hematological Cancers: Present and Future". In: *HemaSphere* 5.2 (2021), e540. DOI: 10.1097/HS9.0000000000000540.
- [10] Intelligen, Inc. *Accessing SuperPro Designer Variables with COM*. https://docs.intelligen.com/spd/SPDWebHelp/index.htm#Bk7/AppendixE/Accessing_Pro_Designer_Variables_with_COM.htm#XREF_66613_E_4_Accessing_Pro. SuperPro Designer Documentation, accessed December 2025. 2024.
- [11] S. S. Farid et al. "A systematic framework for the techno-economic evaluation of cell therapy manufacturing". In: *Biotechnology and Bioengineering* 111.5 (2014), pp. 1029–1040. DOI: 10.1002/bit.25162.
- [12] A. S. Simaria et al. "Allogeneic cell therapy bioprocess economics and optimization". In: *Biotechnology and Bioengineering* 111.1 (2014), pp. 69–83. DOI: 10.1002/bit.25008.
- [13] Richard Turton et al. *Analysis, Synthesis, and Design of Chemical Processes*. 5th ed. Boston, MA: Pearson Education, 2018.
- [14] Andrea Saltelli et al. *Global Sensitivity Analysis: The Primer*. Chichester, UK: John Wiley & Sons, 2008.
- [15] Francesca Pianosi et al. "Sensitivity analysis of environmental models: A systematic review with practical workflow". In: *Environmental Modelling & Software* 79 (2016), pp. 214–232. DOI: 10.1016/j.envsoft.2016.02.008.
- [16] Andrea Saltelli et al. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index". In: *Computer Physics Communications* 181.2 (2010), pp. 259–270. DOI: 10.1016/j.cpc.2009.09.018.
- [17] I. M. Sobol. "Global sensitivity indices for non-linear mathematical models and their Monte Carlo estimates". In: *Mathematics and Computers in Simulation* 55.1–3 (2001), pp. 271–280. DOI: 10.1016/S0378-4754(00)00270-6.
- [18] Andrea Saltelli. "Global sensitivity analysis: An introduction". In: *Proceedings of the 4th international conference on sensitivity analysis of model output (SAMO 2004)* (Jan. 2004).
- [19] Jonathan Herman and William Usher. "SALib: An open-source Python library for sensitivity analysis". In: *Journal of Open Source Software* 2.9 (2017), p. 97. DOI: 10.21105/joss.00097.
- [20] Nestor V. Queipo et al. "Surrogate-based analysis and optimization". In: *Progress in Aerospace Sciences* 41.1 (2005), pp. 1–28. DOI: 10.1016/j.paerosci.2005.02.001.
- [21] Alexander I. J. Forrester, Andras Sobester, and Andy J. Keane. *Engineering Design via Surrogate Modelling: A Practical Guide*. Chichester, UK: John Wiley & Sons, 2006.
- [22] Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach*. 4th ed. Mason, OH: South-Western Cengage Learning, 2009.
- [23] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York, NY: Springer, 2009.
- [24] Raymond H. Myers, Douglas C. Montgomery, and Christine M. Anderson-Cook. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. 4th ed. Hoboken, NJ: John Wiley & Sons, 2016.
- [25] Kun Cheng, Wei Zhang, and Yi Li. "A review of surrogate modeling in engineering design". In: *Structural and Multidisciplinary Optimization* 62 (2020), pp. 119–142. DOI: 10.1007/s00158-019-02450-4.
- [26] Antonio Del Rio Chanana. *Lecture 1: Regression*. ML4CE lecture notes, Imperial College London. 2024.
- [27] Sylvain Arlot and Alain Celisse. "A survey of cross validation procedures for model selection". In: *Statistics Surveys* 4 (2010), pp. 40–79. DOI: 10.1214/09-SS054.

Author Index

Linked to paper number

Abdullahi, Amiira	43	Hoyle, Alfred	1
Adan, Adam	13	Huang, Eric Sihan	7
Adiraputri, Alya Dania	52	Huang, Shaozhi	26
Adjetey, Samantha	3	Huang, Zhixiang	11
Ahmed, Anam	57	Jadav, Sanjiv	14
Akgök, Eylül	20	Jasani, Shaylan	8
Al-Aedroos, Hamza	47	Jayawant, Eric	27
Amuwo, Keziah	45	Jeyavathanan, Amirthan	50
Ashraf, Awaad	34	Jia, Angelina	58
Azmi, Ammar	38	Jindal, Aarna	18
Bajaj, Koyna	18	Joshua, Michael	8
Barbacci-Schettino, Marco	48	Kashyap, Ayushman	59
Berri, Lynn	42	Keomany, Uma	62
Blasco, Enrique	51	Kholwadiab, Idrees	54
Budge, Faith	41	Knoyle, Katy	64
Bull, Daniel	46	Koçan, Hakan	10
Burke, Jack	65	Lam, Tak Fai	17
Butler, Loris	41	Lauhabutr, Kasidit	16
Chatha, Gilby	44	Lee, Helen	15
Chen, Mingze	33	Li, Yatian	23
Chen, Xin	9	Li, Yi	9
Cheng, Zhengxi	33	Liu, Mason	26
Cheong, Cheng Xuan	38	Liu, Shuojun	40
Chern, Jaace	4	Liu, Sisi	64
Chew, Alexa	2	Lloyd, Ross	27
Chin, Wei Jen	56	Lun, Eric	28
Choithani, Shaurya	22	Mahitkumar, Kundan	54
Dandjee, Nora	30	McGrath, Isabella	25
De Sousa Trad, Maria Fernanda	6	Miah, Mohammed Nadim	63
Drury, William	5	Middleton, Robert	59
Fang, Zoe	66	Mitra, Julia	51
Fayad, Rebecca	69	Mollatt, Zachariah Alan	6
Gadhia, Simran	31	Na, Tony	46
Gan, Xuan	36	Oh, Minsun	45
Gao, Walt	61	Okoye, Onyi	3
García Belza, Marta	69	Paiboonsirijit, Katayarn	68
Gibilaro, Lucia	25	Parekh, Shivani	43
Gong, Fanqi	49	Piao, Xingyu	19
Gong, Xiquan	36	Poomore, Maungo Romeo	56
Gu, Xinyuan	40	Ponsford, Joseph	67
Guo, Sisi	39	Pradadeville, Gabriel	24
Gyuris, Eszter	37	Qian, Yu	21
Hayat, Ali	42	Rafiq, Aliff	13
Ha, Rachel	66	Richards, Erin	37
He, Jiayi	32	Rigby, Rachel	52
Ho, Venus	55	Rogers, Thomas	16

Rong, Haotian	11
Sajed, Daanyaal	7
Salame, Rami	50
Salastekar, Aryan	22
Sarker, Abid	63
Sivakumar, Shreyas	47
Sivanehru, Kathiir	31
Slade, Thomas	48
Soo, Jing Qiao	52
Sturina, Samanta	68
Tabassum, Nazifa	57
Tang, Angel	55
Tang, Winston	10
Thill, Alexia	58
Tholley, Nadia	62
Too Kah Fung, Justin	17
Tung, Callum Yuan	24
Turányi, Vilmos	5
Turner, Alexander	39
Tewinpagti, Bam	2
Wan Rawi, Wan Iqbal	4
Waraich, Hassaan	34

Weber, Quiterie	30
Weston, Christopher	12
Williams, Tal	12
Wood, Joe	1
Wright, Joseph	52
Wu, Tong	34
Wu, You	61
Xu, Xinyue	21
Yan, Chenyu	23
Yastrib, Maya	44
Yoshizaki, Yuta	14
Zeng, Adam	49
Zhang, Candy	19
Zhang, Yuye	60
Zhao, Tingxu	60
Zhao, Qianyi	32
Zhao, Ziheng	29
Zhu, Alice	28
Zhu, Liguang	34

Supervisor Index

Linked to paper number

Adjiman, Claire	22, 25
Cabral, João	6, 57, 59
Chachuat, Benoit	38, 65, 66
Chen, Rongjun	32, 36, 56
Ceroni, Francesca	68
del Rio Chanona, Antonio	3, 31
Elany, Yuval	20
Eslava, Salvador	19, 44, 60
Fennell, Paul	49, 53, 67
Hallett, Jason	15, 41, 62
Hankin, Anna	9, 52, 55
Hawkes, Adam	8, 17, 18
Hellgardt, Klaus	12, 45
Heng, Jerry	13, 46, 64
Jackson, George	1, 7
Li, Kang	21, 39
Matar, Omar	47, 48
Mercangoz, Mehmet	5, 29
Müller, Erich	42, 63
Petit, Camille	11, 23, 43
Pini, Ronny	27, 28, 37
Papathanasiou, Maria	33, 54, 69
Shah, Nilay	10, 16, 34, 50
Shah, Umang	14
Song, Qilei	16, 34
Sormanni, Pietro	58
Titirici, Magda	35, 40, 51
Trusler, Martin	4
Williams, Daryl	61
Xu, Yun	26, 30
Yetisen, Ali	2, 24

