

IMPERIAL

On-chip III-V Semiconductor Network Lasers for Neuromorphic Computing

A thesis submitted in partial fulfilment of the requirements for the PhD degree.

Jakub Bartosz Dranczewski

Imperial College London

Department of Physics

June 2025

Supervision:

Prof. Riccardo Sapienza

Imperial College London

Prof. Kirsten Moselund

Paul Scherrer Institut,

EPFL

Heinz Schmid

IBM Research Europe – Zürich

Abstract

Modern Artificial Intelligence workloads are placing a growing amount of strain on digital hardware, increasing costs and energy demands. The field of neuromorphic computing seeks to help alleviate this problem through physical platforms that could be used as accelerators for machine learning tasks. Network lasers are a promising but unexplored platform for this application – they support a high number of interacting lasing modes through multiple scattering between defined waveguiding paths, and their complexity, non-linearity, sensitivity, and compact size address several challenges facing the field of photonic neuromorphic computing. To enable their use as a computing platform, we need to move them from the current dye-doped electrospun polymer platform to a lithography-defined semiconductor process, which allows for fully designable lasers, and high resiliency to degradation under pumping.

To implement this vision, I first explore the etching of III-V semiconductor layers to create photonic nanodevices. Parameter optimisation allows us to produce high quality structures from InP, our main material of interest, as well as other epitaxial platforms like InGaAs quantum wells. Next, I study the lasing from etched microdisk devices, presenting multiple ways in which lasing properties can be used to inform fabrication choices. With fabrication and lasing performance validated, I present experimental measurements from etched semiconductor network lasers, investigating the physics behind the lasing modes and their coupling. Finally, I show the application of network lasers to neuromorphic computing, first through edge detection, which matches the mode competition dynamics of the laser, and then for image classification, also combining the two schemes to form a photonic convolutional neural network. We obtain high classification accuracies on tasks including MNIST, Fashion MNIST, and BreA KHis 400X, especially in the few-shot regime, and explore the likely physical basis of the neuromorphic performance.

Copyright and originality statements

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC 4.0).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that you credit the author and do not use it, or any derivative works, for a commercial purpose.

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. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

The text above is a summary and does not constitute the legally-binding license. Please consult the complete text of the CC BY-NC 4.0 license at <https://creativecommons.org/licenses/by-nc/4.0/legalcode.en> for the full details of this agreement.

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.

The work, text, and figures presented in this thesis are my own unless otherwise stated. Where work from publications on which I am co-author is adapted, this is made clear in the heading of the corresponding chapter.

Acknowledgments

“Journey before destination.”

Some may call it a simple platitude, but it is far more. A journey will have pain and failure. It is not only the steps forward that we must accept. It is the stumbles. The trials. The knowledge that we will fail. [...] But if we stop, if we accept the person we are when we fall, the journey ends. That failure becomes our destination. To love the journey is to accept no such end. I have found, through painful experience, that the most important step a person can take is always the next one.

Oathbringer, BRANDON SANDERSON

A PhD is a journey, and I want to start its final act by thanking those who have helped me along the way. This PhD was also a journey in the more literal sense, from Zürich to London, and there are people in both places that have supported me through it and made the experience truly special.

First, I want to thank my supervisors, Prof. Riccardo Sapienza, Prof. Kirsten Moselund, and Heinz Schmid, who have not only been incredible guides on my first foray into fabrication and complex lasing, but have also built the brilliant teams I joined and felt immediately welcome in. Kirsten, diving into fabrication for the first time was daunting, but I always looked forward to your weekly guidance and enthusiasm, broadening my horizons significantly. Heinz, thank you for your immense technical expertise, and always asking just the right questions about any problem. Riccardo, your careful headhunting and vision over six years ago is what brought me into the field of nanophotonics, I could not have dreamed of a better mentor and collaborator in the pursuit of knowledge. A special thank you also to Jack Gartside, with whom I started working closely when I came back to Imperial, and whose vast neuromorphic computing insight and careful guidance have helped me spread my wings in a new field.

IBM Research Europe – Zürich was a new place for me, and I am thankful to their team, and my manager Heike Riel, for enabling the partnership of the CORAL EU project. I also thank the amazing team of collaborators and friends I met there, first in the MIND team, and then in QTP and the III-V sub-team – the fact that there is too many names to list here is a testament to the wonderfully tight-knit community you have built, which truly made Zürich feel like a new home to me. I will single out Markus, whose unwavering friendship and competence were always a source of inspiration, and Isa, without fail an amazing office and cleanroom conversation partner. Anna Fischer was my project partner throughout CORAL, but I got to work with her most closely during the months of overlap at IBM, and so I thank her here. You were an amazing project partner, and I am extremely grateful for your support in everything. You have done great things in this project, and no doubt are going to do great things after. And on the technical end, I also thank the Cleanroom Operations Team of the Binnig and Rohrer Nanotechnology Center (BRNC) for their help and support, and the HR team for their patience and assistance.

In London, Riccardo’s team has made my PhD time similarly amazing. A special thank you goes to Kit and Dhruv, who have played a huge part in establishing the measurement system, and especially Kit for being the first other user of many of my weirder measurement scripts, and showing me how to get around the setup. On the technical side, a huge thank you to Stefano as well, who has always been the backbone of the experimental work. Raziman, thank you for trying to explain theory to me whenever I felt like sticking my head out of the lab. And Elias, Paul, and Tobias, thank you for exploring the more computational side of the project, you were a joy to co-supervise. On the less technical side, much love to the other inhabitants of the H722 office for the many conversations and lunches, especially Anthony for sharing lab time with us, and listening to my many rants.

Closing up on the personal side, I want to thank my family and friends, who have been a steady support for me throughout these turbulent years. Developing a chronic condition (long COVID) in the first year of one’s PhD is not easy to navigate, both on the professional and private fronts, and without your support I do not believe I would have managed. So first, I thank my family in Poland, who have championed and nurtured my interests since the very beginning – you are a huge reason for my curiosity and drive. Thank you from the bottom of my heart for your unending love and support. Then, my friends, in roughly chronological order: those in Poland, who have been with me from the very start, and make sure I do not forget the best language in the world; those in London, who were the first to welcome me in this home away from home; those on Discord, who kept me connected as I found my place in a new country,

ACKNOWLEDGMENTS

and continue to always be there for me; those in Zürich, with whom I share many a story of Zürichstrasse and beyond; and those in the Imperial College Dramatic Society, for welcoming an old soul back into the theatrical fold.

You have all helped me along this journey, and as I reach this particular finish line, I am most excited to see what the next steps will be.

List of publications

Major contributions

- **J. Dranczewski**, A. Fischer, P. Tiwari, M. Scherrer, D. Saxena, H. Schmid, R. Sapienza, K. Moselund, “[Plasma etching for fabrication of complex nanophotonic lasers from bonded InP semiconductor layers](#)” *Micro and Nano Engineering*, vol. 19, p. 100196, Jun. 2023.
- D. Saxena,* A. Fischer,* **J. Dranczewski**, W. K. Ng, N. V. Trivino, H. Schmid, T. V. Raziman, A. Arnaudon, M. Barahona, O. Hess, K. Moselund, R. Sapienza, “[Designed Semiconductor Network Random Lasers](#),” *Laser & Photonics Reviews*, vol. 19, no. 1, p. 2400623, 2024.
- **J. Dranczewski**,* W. K. Ng,* A. Fischer,* T. V. Raziman, D. Saxena, T. Farchy, K. Stenning, J. Peters, H. Schmid, W. R. Branford, M. Barahona, K. Moselund, R. Sapienza, J. C. Gartside, “[Retinomorphic Machine Vision in a Network Laser](#),” *arXiv*, Aug. 2025. *[manuscript in review]*

Minor contributions

- A. Fischer, T. V. Raziman, W. K. Ng, J. Clarysse, D. Saxena, **J. Dranczewski**, S. Vezzoli, H. Schmid, K. Moselund, R. Sapienza, “[Controlling lasing around exceptional points in coupled nanolasers](#),” *npj Nanophotonics*, vol. 1, no. 1, pp. 1–6, May 2024.
- A. Fischer,* T. Severs Millard,* X. Xiao, T. V. Raziman, **J. Dranczewski**, R. C. Schofield, H. Schmid, K. Moselund, R. Sapienza, R. F. Oulton, “[Surface Lattice Resonance Lasers with Epitaxial InP Gain Medium](#),” *ACS Photonics*, vol. 11, no. 10, pp. 4316–4322, Oct. 2024.

*Authors contributed equally.

Contents

Title page	1
Abstract	2
Copyright and originality statements	3
Acknowledgments	4
List of publications	7
Contents	8
List of Figures	11
Acronyms	14
1 Introduction	16
1.1 Challenges in modern computing and machine learning	16
1.2 Neuromorphic computing	20
1.3 Network lasers as a platform for neuromorphic computing	22
1.4 Thesis outline	25
2 III-V fabrication and optimisation	27
2.1 III-V semiconductors and devices	28
2.2 Direct Wafer Bonding for III-V integration	29
2.3 Principles of plasma etching	31
2.4 Exploring etch recipes	34
2.4.1 Chlorine-based etching	35

2.4.2	Methane-based cyclic etching	35
2.5	Etching Optimisation	37
2.5.1	Design of Experiment principle	37
2.5.2	Methane-based continuous etching	38
2.6	Case study: InGaAs Quantum Wells	40
3	Semiconductor lasers for fabrication validation	43
3.1	Introduction to lasing	44
3.2	InP microdisk lasers	45
3.3	Experimental studies of lasing	47
3.4	Lasing threshold statistics for fabrication evaluation	49
3.4.1	Lasing from InP microdisks for different etching chemistries	51
3.4.2	Quantum well lasing results	52
3.4.3	Passivation and its impacts on cavity degradation	53
3.5	Other lasing-based diagnostics	55
4	Semiconductor network lasers	57
4.1	System introduction	58
4.2	The Physics of network lasers	62
4.2.1	Transport within the network and coupling between the waveguides	62
4.2.2	Lasing modes within a random laser	65
4.2.3	Lasing mode competition	68
4.2.4	Long-distance coupling and disturbing it	70
4.3	Application considerations	71
4.3.1	Sensitivity and resilience	71
4.3.2	Network design and topology	72
5	Neuromorphic computing with network lasers	75
5.1	Feature detection	76
5.1.1	Lasing mode competition for convolution-like feature detection	76
5.1.2	The 4x4 image library	80
5.2	Extreme learning machines – image classification	84
5.3	Photonic Convolutional Neural Network	87
5.4	Harder tasks and few-shot regimes	88

5.5	Physical basis for neuromorphic performance	92
6	Conclusions	95
6.1	Summary	95
6.2	Outlook and future work	97
6.2.1	Energy use and speed	97
6.2.2	Material and fabrication	98
6.2.3	Tuning and network topology	98
6.2.4	Segmentation and other novel tasks	100
A	Experiment automation – puzzlepiece	102
B	Additional figures	105
C	Fabrication images	114
	Reproduction permissions	120
	Bibliography	122

List of Figures

1.1	Facets of Moore's Law	17
1.2	Classical computing and machine learning fundamentals	19
1.3	AI and the memory wall	20
1.4	The two main classes of neuromorphic architectures	21
1.5	Random network lasers	23
1.6	Network laser as a physical data processor	25
1.7	Thesis outline	26
2.1	Direct wafer bonding	29
2.2	Lithography for mask creation	30
2.3	Etch diagnostic tools	31
2.4	Fabrication and diagnostic tools	32
2.5	Boiling points for III-V semiconductor etch products	33
2.6	Cl ₂ plasma etching	35
2.7	Cyclic CH ₄ plasma etching	36
2.8	Plackett-Burman factorial DoE principle	38
2.9	DoE etching study results	39
2.10	Continuous CH ₄ plasma etching	40
2.11	Quantum well etching	41
2.12	The optimised continuous CH ₄ etch chemistry	42
3.1	The core principles of lasing	45
3.2	Confinement and amplification in an InP microdisk laser	46
3.3	Diagram of the optical setup	47
3.4	Lasing spectrum from an InP microdisk	49

LIST OF FIGURES

3.5	Disks with varying radii cover the material's gain spectrum	50
3.6	Lasing thresholds measured for disks etched with different chemistries	51
3.7	Quantum well lasing results	53
3.8	The impact of passivation on device degradation	54
3.9	Other potential lasing-based fabrication diagnostics	56
4.1	InP semiconductor lasing networks overview	59
4.2	Diagram of the second optical setup	60
4.3	Exploring network laser spectra	61
4.4	Studying light transport within the networks	63
4.5	Waveguiding and waveguide coupling within the network	64
4.6	Selected simulated network mode profiles	66
4.7	Selected measured network mode profiles	67
4.8	Non-linear behaviour of modes in a network laser	69
4.9	The principle of mode competition	70
4.10	Losses disturbing long-range coupling in a network laser	71
4.11	Spectral sensitivity and emission stability of InP network lasers	72
4.12	Lasing emission from four examples of network topologies	73
5.1	Convolutional processing through a complex network laser	77
5.2	Exploring feature detection in detail	79
5.3	RGB edge detection across raster window dimensions	80
5.4	The 4x4 image library	81
5.5	Feature map sweeps for modes 0-9 from the 4x4 library	83
5.6	Average lasing hyperspectra for the 10 classes of MNIST digits	84
5.7	Single layer classification architecture	86
5.8	Two layer architecture – a Photonic Convolutional Neural Network	87
5.9	Classification results for Fashion MNIST	89
5.10	Classification results for BreKHis 400X	90
5.11	Accuracy vs training length for our physical system and fully digital solutions	91
6.1	Pre- and post-fabrication tuning opportunities	99
6.2	Network lasers for image segmentation	100
A.1	Experiment automation with puzzlepice	103

B.1	Microdisk lasing threshold dependence on pump position	105
B.2	Network lasing with picosecond pump laser	106
B.3	Simulated waveguide mode profiles and effective refractive indices	106
B.4	Black and white cameraman image	107
B.5	Mode and feature maps for all library modes	108
B.6	Linear fitting to library mode intensities	109
B.7	Example edge maps for MNIST	110
B.8	Example edge maps for Fashion MNIST	111
B.9	Logistic regression weights for a two layer classifier	112
B.10	Number of features selected in hyperparameter optimisation	113
C.1	Cracked optical resist	114
C.2	SiO ₂ deposited on optical mask	115
C.3	Wax bubbles on ICP carrier wafer	115
C.4	Lichtenberg figures on the ICP carrier wafer	116
C.5	E-beam patterns on a large disk pattern	116
C.6	Polymer nano-spheres after plasma etching	117
C.7	Nano-spheres clustering around debris	117
C.8	Poorly cleaved chip edge with an optical mask	118
C.9	Failed etch test	118
C.10	Material remnants between waveguides	119
C.11	Side view of an etched network laser	119

Acronyms

Computing

AI	Artificial Intelligence
API	Application Programming Interface
CNN	Convolutional Neural Network
CPU	Central Processing Unit
GPU	graphics processing unit
GUI	Graphical User Interface
LLM	Large Language Model
ML	Machine Learning
MLP	multilayer perceptron
RAM	Random Access Memory
RGB	red, green, and blue

Fabrication

ALD	Atomic Layer Deposition
BOX	buried oxide
DoE	Design of Experiment
DWB	Direct Wafer Bonding
FIB	Focused Ion Beam
HSQ	hydrogen silsesquioxane
ICP	Inductively Coupled Plasma

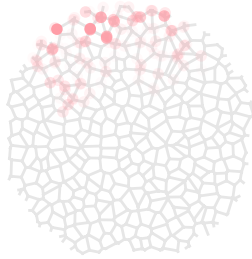
MOCVD	metalorganic chemical vapour deposition
PECVD	plasma-enhanced chemical vapour deposition
QD	quantum dot
QW	quantum well
RF	Radio Frequency
RIE	Reactive Ion Etching
TASE	Template-Assisted Selective Epitaxy
VCSEL	Vertical Cavity Surface Emitting Laser

Measurements & Equipment

CCD	charge-coupled device
CW	continuous wave
DMD	Digital Micromirror Device
LL	light-in light-out
ND	neutral density
OPA	Optical Parametric Amplifier
PL	photoluminescence
SEM	Scanning Electron Microscope
SPD	Single Photon Detector

Theory & Simulations

FDE	Finite Difference Eigenmode
FDTD	Finite-Difference Time-Domain
SALT	steady-state ab initio
TE	Transverse Electric
TM	Transverse Magnetic
TIR	Total Internal Reflection
WGM	Whispering Gallery Mode



Chapter 1

Introduction

Computing has taken the world by storm. Most people own a number of ‘smart’ devices: computers, laptops, phones, game consoles, and increasingly home appliances. As these grow smaller, more portable, and more powerful, one may ask what challenges remain in the field of computation, and what is the next step after the wildly successful age of the silicon transistor. In this thesis I will explore one potential approach to the next chapter of computing – utilising physical systems with dynamics that match the computation we want them to perform. But first, let us explore the problems presently facing classical computing, and how these are changing with the proliferation of artificial intelligence.

1.1 Challenges in modern computing and machine learning

Moore’s Law is an observation and prediction originally made by Gordon Moore in 1965^[1] and revised in 1975,^[2,3] noting that the number of transistors in an integrated circuit was doubling roughly every two years, and it would likely continue to do so for some time. Many have proclaimed the death of Moore’s Law since,^[3] as transistor sizes start to approach physical limits: maximum dopant density, electric breakdown of the thinning insulator layers, electron tunnelling,^[4] as well as thermal dissipation issues.^[5] The reports of its death are fortunately somewhat exaggerated, as breakthroughs in semiconductor fabrication and transistor gate shape have allowed further minification, with other techniques proposed to keep increasing transistor density.^[6] Data from Ref. [7] adapted in Fig. 1.1 a shows that the number of transistors in microprocessors has indeed kept increasing at a steady, logarithmic pace.

Looking beyond the initial formulation of transistor number, we can see scaling issues arising a

bit more clearly though. Fig. 1.1 a also shows the clock frequency of microprocessors, and a figure of merit for single-thread performance. Both of these have plateaued in the early 2000s. Similarly, the energy required for a switching operation in transistors continues to decrease, but the pace of this has slowed since the turn of the century, as shown in Fig. 1.1 b. The complexity and scale of our computational systems is increasing, but the progress in their individual components does appear to be slowing.

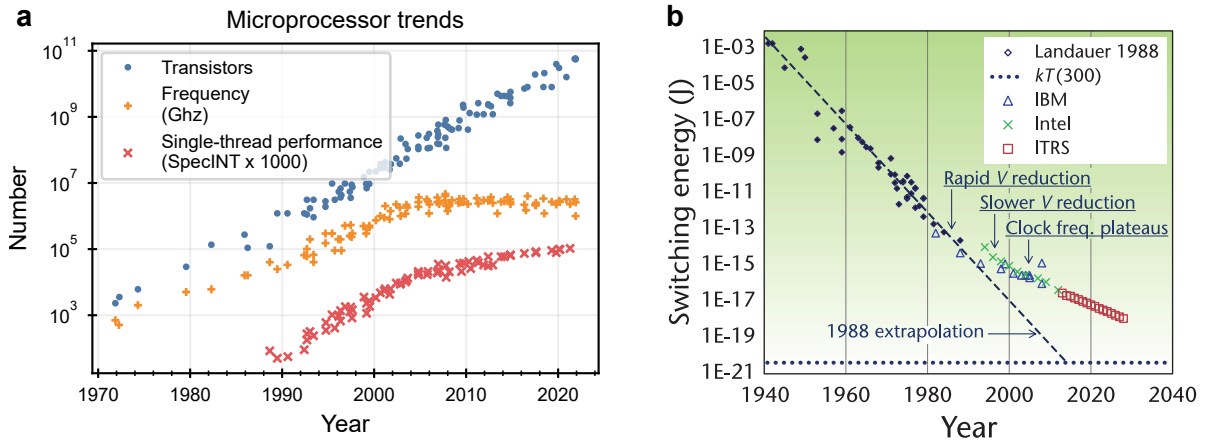


FIGURE 1.1: Facets of Moore’s Law. **a** The number of transistors in microprocessors has continued to grow exponentially as predicted, but the clock frequencies and single-thread performance have largely plateaued. Data adapted from Ref. [7] – CC BY 4.0. **b** The switching energy of transistors has been decreasing exponentially, but this trend has slowed down in the past decades. Adapted from Ref. [8] – © 2017 IEEE.

These growing limitations become increasingly apparent as we demand more and more from our compute systems. The recent raise of *Large Language Models* (LLMs) and video-processing models is putting increased strain on data centre architecture, with the number of parameters in LLMs in particular growing on average 410x times every two years,^[9] and the amount of compute used in training the models doubling every 3.4 months, outpacing Moore’s Law’s growth. Model performance generally scales with the parameter number,^[10] and so the compute demand is likely to continue growing. With more hardware required to match said demand, the energy needs of *Artificial Intelligence* (AI) are also growing. The exact impact is difficult to predict, but data centre electricity use is predicted to double by 2030,^[11] accounting for 7% to 15% of electricity use in the United States by 2030, with specific areas impacted more as data centres tend to be built close to each other.^[12] To understand this problem and the proposed solutions better, let us look at some fundamental principles of classical computing.

Most computers used nowadays are based on the von Neumann architecture,^[13,14] illustrated in Fig. 1.2 a. An abstracted ‘computer’ consists of a *Central Processing Unit* (CPU) and a

memory unit (storing both data and instructions). The CPU has a connection to the memory unit, over which instructions and data are transmitted to be processed by the CPU's Control Unit (which orchestrates the operations to be performed), and the Logic Unit (which implements basic arithmetic calculations). The memory is usually split into *Random Access Memory* (RAM) and caches, which are closer to the CPU and fast, but limited in size, and disk storage, which can be much greater in size, but slower to get to and operate.

The need to shuttle instructions and data between the CPU and the memory unit is a major constraint of this architecture.^[15] The energy cost of data transmission is actually much greater per bit than switching a transistor gate, and does not scale in line with our advancements in transistor engineering.^[16] This is most apparent for computing problems that require many distinct mathematical operations to be performed on varying data – the computer cannot keep all the values in a CPU cache and has to perform many round-trips to the RAM

An example of such a computing problem is a neural network, illustrated in Fig. 1.2 b. Neural networks are constructed out of digital ‘neurons’ (orange in the diagram), sometimes referred to as ‘perceptrons’. These are meant to emulate the way neurons in a brain are triggered to send signals on. They sum the inputs provided using a matching number of weights, apply an optional biasing value, and then pass the sum through a non-linear ‘activation’ function, which can for example set negative values to zero.^[17] This style of network is referred to as a *multilayer perceptron* (MLP). It is usually trained with backpropagation, where the output layer is compared to the desired answer, and the difference between these is fed back through the network, changing the weights such that the output is closer to being right. This is repeated with many training examples, until the network is good enough to provide correct answers for data it has not seen before.

Machine vision, where tasks such as feature detection, classification, and segmentation are performed on image data, represents a significant portion of modern machine learning demand.^[17,18] *Convolutional Neural Networks* (CNNs) are an architecture widely used in machine vision applications.^[19,20] They begin with feature detection layers, where multiple convolutional kernel filters are separately raster-scanned across an image, performing multiplications between parts of the image and the kernel. This operation is illustrated in Fig. 1.2 c. The results of these multiplications are assembled into ‘feature maps’ of the input image, with different kernels detecting features such as edges, corners, and curves. The feature maps are then fed into more typical neural network layers, forming the CNN.

The convolutional feature detection greatly enhances machine vision performance compared

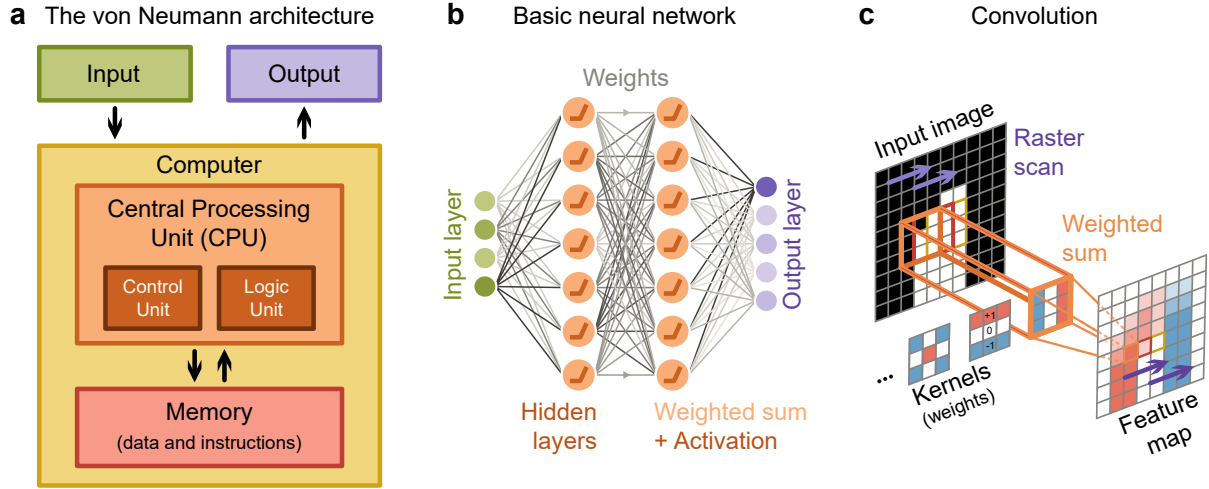


FIGURE 1.2: Classical computing and machine learning fundamentals. **a** Most modern computers are based on the von Neumann architecture. This diagram illustrates its main components. **b** A *multilayer perceptron* (MLP) – the fundamental neural network design. It consists of multiple interconnected layers, with data flowing left to right on the diagram. **c** Convolution for feature detection. Small windows of the image are multiplied by a selected kernel, with the result highlighting positive (red) and negative (blue) edges.

to traditional MLP neural networks^[17–20] as it allows the neurons to be more spatially aware and ‘see’ the neighbourhood of a specific pixel, but is fairly computationally intense.^[21]

In general, each layer of a neural network is comparable to a high-dimensional vector-matrix multiplication, with many weights and values which have to be moved between the logic unit in the CPU and the RAM. This type of computation is therefore inherently mismatched to the underlying von Neumann architecture, and the moving of data accounts for much of the total energy used to run an AI model.^[22] *Graphics processing units* (GPUs) alleviate this somewhat by doing many multiplication operations in parallel, which minimises the time spent transferring data, but current hardware is still limited – Fig. 1.3 shows that the number of parameters in neural networks is growing much faster than the amount of available RAM in GPU-style accelerators. This means that the model weights have to be transferred back and forth from the even slower drive storage, massively escalating the time and energy taken, especially when training large models.^[9]

The growing demands of modern AI are exposing a fundamental mismatch between the architectures we use for computing and for neural networks. Computers are linear and digital (using binary numbers to represent data), but neural networks are analogue and non-linear, and their sheer size clashes with the memory-processing separation of von Neumann computers. Our brains, which neural networks are inspired by, match these requirements better, learn quickly,^[24]

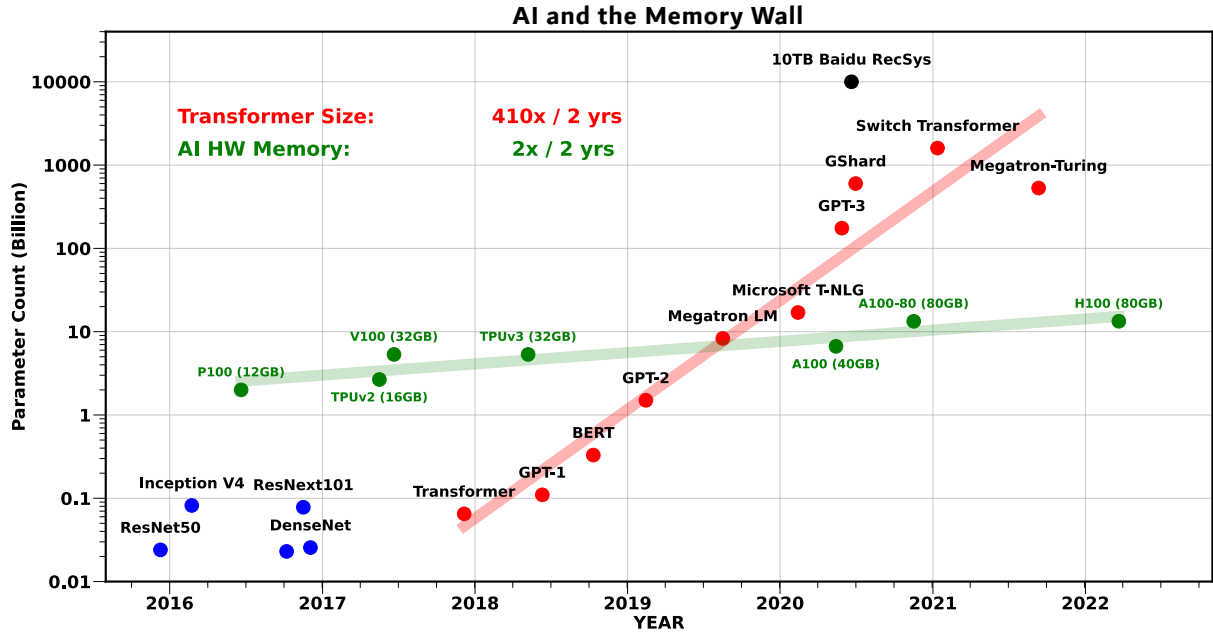


FIGURE 1.3: The number of parameters used by LLMs has been increasing exponentially, but the amount of RAM available in AI acceleration hardware has not been keeping up. If a model cannot be kept fully in RAM, its weights need to be moved back and forth between the processing unit and memory drives, decreasing efficiency. Adapted from Ref. [9, 23].

and are extremely efficient, using only ~ 20 W,^[25] compared to hundreds of Watts used by modern GPUs. There is therefore much interest in extending the similarities to the brain’s operation from the software to the compute hardware.

1.2 Neuromorphic computing

The field of making computational hardware more similar to the brain (‘neuromorphic’) is a broad one. Fig. 1.4 illustrates the two main divisions of it.

‘Isomorphic’ systems implement the operations characteristic of digital neural networks as discussed in Section 1.1, for example matrix multiplication.^[26] They resemble the brain in their departure from the von Neumann architecture, bringing memory and computation into the same location, and letting them interact more closely. They also aim to allow parallelism and acceleration through techniques like wavelength multiplexing and the inherently fast timescales of many physical phenomena – for example, combinatorial problems are hard to solve digitally, but can be mapped to physical Ising machines, which quickly settle into an equilibrium corresponding to the solution.^[27] Various physical computing architectures take inspirations from aspects of how neurons communicate,^[28] for example working with spike signals,^[29, 30] letting the data modify the neuron to form long- and short-term ‘memory’,^[31] or synchronising the oscillations

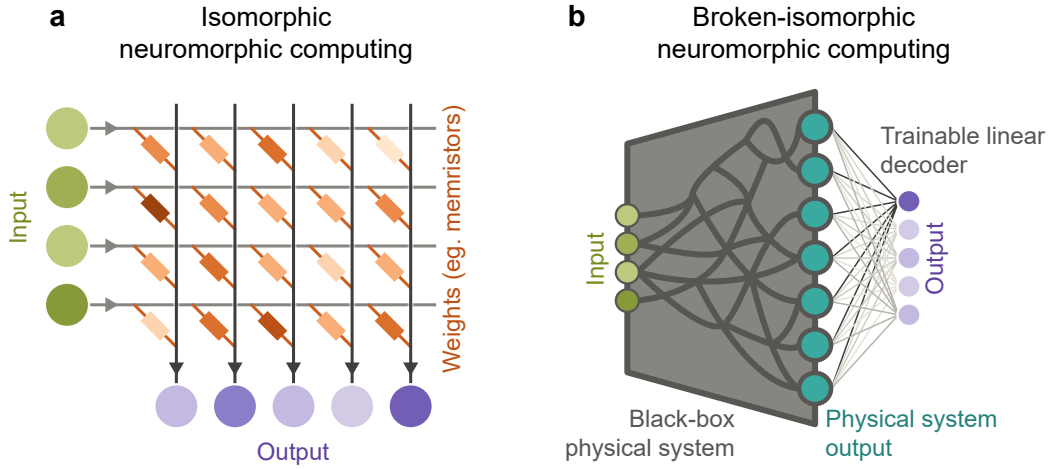


FIGURE 1.4: Examples of the two main classes of neuromorphic architectures. **a** Isomorphic systems aim to be a direct accelerator for machine learning operations like matrix multiplication. Here, a cross-bar array of memristors is shown. Their resistances act as weights, dividing the input voltage across the vertical output bars. **b** Broken-isomorphic systems act as complex data processors, not trying to implement a specific digital architecture. A trainable layer after the system translates its output into the desired format. This is usually digital, but can in theory be an isomorphic physical system.

of multiple ‘neurons’ together.^[32] Specific properties of the physical neurons can also enable brain-like training algorithms,^[26] though the neurons imprecision can hinder more traditional approaches like backpropagation.^[28]

Examples of isomorphic neuromorphic systems include cross-bar arrays of memristors (whose resistance can be changed by the application of specific currents),^[33–35] free-space optics utilising absorption,^[36–38] phase shift,^[39,40] or metalenses^[41] to provide weights, and on-chip photonics employing either Mach-Zehnder interferometers^[42,43] or phase-change materials for absorption.^[44]

‘Broken-isomorphic’ systems do not try to directly resemble a neural network, but rather rely on complex, non-linear interactions within physical objects to encode data into high-dimensional spaces that may be useful for computation.^[26] This usually necessitates a simple trainable decoder at the end to produce useful ‘answers’ out of the system’s output, which is often akin to a fully linear, single-layer version of the [MLP](#) pictured in Fig. 1.2 **b**. Most of the ‘computation’ happens in the physical system (making a complex problem linearly separable), with the number of parameters in the digital layer much smaller than for a fully classical approach,^[45] addressing the memory and scaling issues of matrix-multiplication-based methods. I discuss the details of how this decoding layer is implemented in Section 5.2.

The two main sub-divisions of broken-isomorphic systems are reservoir computing and extreme learning machines. In reservoir computing the physical system has a degree of memory, such

that past inputs impact current computation, which is most appropriate for temporal tasks like time-series data prediction and transformation,^[46] and speech recognition.^[47] This memory is occasionally achieved through feeding a memory-less system’s output back in with a time-delayed feedback loop.^[48,49] Completely memory-less systems are used in extreme learning machines, where each data transformation is independent from the previous ones, which is suitable for static data tasks^[26] like image classification.^[50] The weak requirements on the properties of the non-isomorphic physical system (as long as it performs a complex, preferably non-linear, transformation of the data) have led to a large variety of published approaches, examples including sound waves, non-linear optics, electronics,^[51] lasers,^[52] spintronics,^[47] and magnets.^[46]

The different physical systems that can be used to implement neuromorphic computing can perform exceptionally well when paired with a task that matches the underlying physics. For example, nanomechanical systems can be tuned to audible frequencies for sound-sensing tasks.^[53] The system’s physics can also limit the suitable applications though. Magnetic systems have excellent memory, but can be slower, and struggle with high-dimensional input. Photonic systems excel at speed and throughput, with high-dimensional inputs achievable through spatial and wavelength multiplexing. On the other hand, they struggle with non-linearity and implementing negative feedback/weights, and do not have memory. Spatial scaling is also generally not favourable, with waveguides taking much more space than electronic transistors, matching light’s longer wavelengths.

In general, a physical system for neuromorphic computing should be sensitive to inputs, map them to complex dynamics, and provide an easily-measurable output. Our research group at Imperial College London thought that we may have a physical system that could be a strong neuromorphic platform, while addressing some of the issues of typical photonic approaches. I will explore its details in the next section.

1.3 Network lasers as a platform for neuromorphic computing

Amplification and confinement can lead to lasing (light amplification by stimulated emission of radiation), and I will explore the details of this process in Section 3.1. But the archetypical laser pointer system of two parallel mirrors is not the only way to confine light – for example, in Section 3.2 I discuss how semiconductor micro-cavities can be used instead. Random lasers are a curious class of laser where the confinement is obtained through multiple scattering,^[54,55] as illustrated in Fig. 1.5 a. This scattering is coherent (which can be seen from speckle patterns

forming when shining a laser pointer at a scattering material), and so some paths within the scattering medium can fulfil phase matching conditions and produce modes, which confine light. With gain (provided either by the scatterers or the medium they are embedded in), these confined modes can lase.

Random lasing has been demonstrated in a variety of media, including powders,^[56] collections of microspheres,^[57] polymer-suspended liquid crystals,^[58] electrochemically-etched semiconductors,^[59] e-beam patterned polymer^[60] or semiconductor layers,^[61] and self-assembling colloidal particles.^[62] The loose requirement on cavity shape, low cost, broad angular emission, complex spectra, and high sensitivity of the lasing modes make random lasers a fitting platform for light-emitting coatings, object identification, and remote sensing.^[54]

The fabrication of most random laser systems is difficult to tune precisely, and simulating multiple scattering from a large number of scatterers is computationally heavy enough to not be practical. In 2019 in Ref. [63] our Imperial College London research group has proposed as system that addresses these issues – random network lasers, as outlined in Fig. 1.5. The system consists of many intersecting dye-doped polymer waveguides fabricated with electrospinning (Fig. 1.5 b). Light scatters between the waveguides at their joining points (‘nodes’), providing

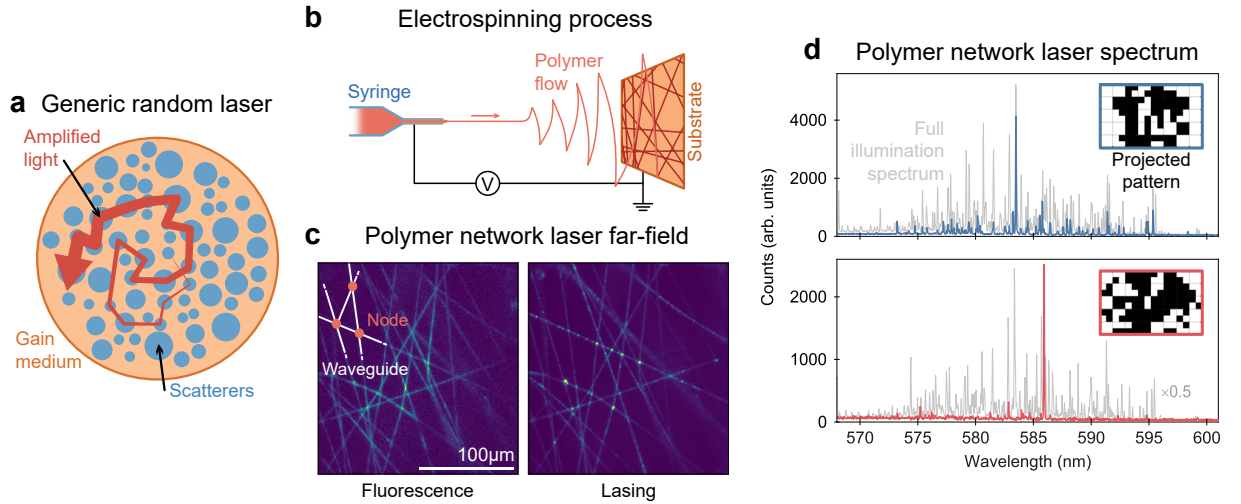


FIGURE 1.5: Random network lasers. **a** The general principle of random laser operation. Scattering provides some confinement, so that light can be amplified by the gain medium of the laser. **b** Electrospinning involves using a high voltage to accelerate material towards the substrate. Electrostatic forces amplify instabilities in flight, resulting in random deposition of continuous fibres,^[64] forming the network laser. **c** Light emission from network lasers created through electrospinning. Randomly oriented waveguides join together at nodes, where light can scatter between them. Below threshold the laser emits uniformly from its waveguides, and above threshold the lasing output is mostly from the nodes. Adapted from Ref. [63] (CC BY 4.0). **d** Sensitivity of the network laser to a projected pattern. By changing which regions of the laser receive gain, one can promote specific modes to lase while suppressing the others. Adapted from Ref. [65] (CC BY 4.0).

the multiple scattering necessary for random lasing, but with both the positions of the scattering points and the paths between them well-defined. This lends itself well to analytical studies, with the network represented as a graph on which Maxwell equations can be solved.^[63]

The system does indeed lase, with Fig. 1.5 c showing the transition between spontaneous emission in random directions below the lasing threshold, to the majority of light being emitted into waveguide modes through stimulated emission above threshold. The lasing spectrum is shown in grey Fig. 1.5 d, consisting of a forest of narrow peaks. Notably, when a pattern is projected onto the laser, its spectrum changes significantly even for small changes in the spatial gain distribution.^[65] The two plots show how the projected pattern can be tuned to obtain nearly single mode lasing. This change is due to a number of processes – the modes are differently localised, with some pump patterns matching their shape better, the shape of the modes changes as the refractive index of the material shifts in illuminated locations, and the modes compete with each other for gain.^[66] I will explore the process of mode competition and its implications in detail in Section 4.2.3.

Random networks lasers are an enticing platform for broken-isomorphic reservoir computing, solving many challenges of conventional photonic systems. They are strongly non-linear, both in the sense that the output increases rapidly around the lasing threshold, and that the many modes are coupled to each other in complex ways. The coupling through mode competition can be suppressive, implementing negative weights. The networks provide a convenient output in the form of the lasing spectrum, which is very sensitive to the projected pattern. They can also have a fairly small footprint of tens of micrometers across, compared to the often millimetre-scale of many optical processors. There was thus interest in testing network lasers as a neuromorphic data processor as illustrated in Fig. 1.6, with the projected image as an input shaping the spatial gain landscape, and the emitted light as the output.

The dye-doped polymer platform unfortunately did not match the neuromorphic ambitions. Photobleaching of the dye causes the performance to degrade over time, quick enough that longer exposure times required to process large machine learning datasets were not practical. The electrospinning process is somewhat tunable, but not reproducible – one cannot make the same network design again, or directly design the network topology.

Direct bandgap semiconductors can provide gain for our lasers, while allowing lithographically defined, reproducible shapes to be etched out. This allows much more design and topology flexibility. Solid-state platforms are also much more resilient to degradation than dye, matching the requirements for the neuromorphic application. They also bring us closer to scalability,

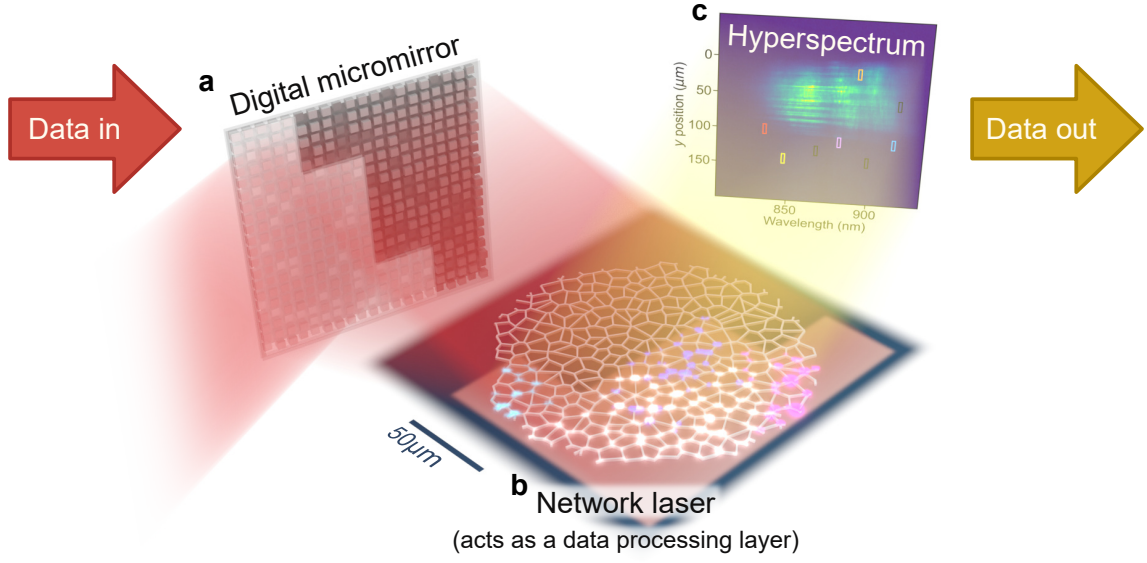


FIGURE 1.6: Network laser as a physical data processor. **a** A *Digital Micromirror Device* (DMD) spatially structures pump light into an arbitrary *input* image, which is illuminated onto an InP network. **b** The pump light provides optical gain in illuminated regions of the InP network, triggering lasing from many different modes at different wavelengths determined by the scattering paths. Spatial mode profiles of three modes are indicated via differently coloured nodes. **c** Lasing spectra emitted by the network are recorded via spatially-resolved spectrometer, forming the *output* of the system. Some spectral regions are highlighted with coloured boxes – these will be relevant for feature detection in Section 5.1. Adapted from Ref. [67].

with semiconductor fabrication processes able to create hundreds of on-chip network devices in parallel. The goal of my work within this thesis is thus to establish a fabrication process for designed semiconductor network lasers, study their physics, and investigate their suitability as a neuromorphic computing platform.

1.4 Thesis outline

Fig. 1.7 shows some key images from each chapter in this thesis. Below, I will provide a brief summary of said chapters.

Chapter 2 describes the process of fabricating III-V semiconductor devices and my attempts to optimise it. I introduce *Direct Wafer Bonding* (DWB) (our integration process of choice), as well as the fundamentals of plasma etching. I demonstrate etching results for InP from a number of chemistries, and show a *Design of Experiment* (DoE) study of how different parameters affect etching outcomes. Finally, I show that the etching process is transferable to other platforms, demonstrating devices made out of InP with embedded InGaAs *quantum wells* (QWs).

Chapter 3 introduces the fundamentals of semiconductor lasing and applies them to fabrication

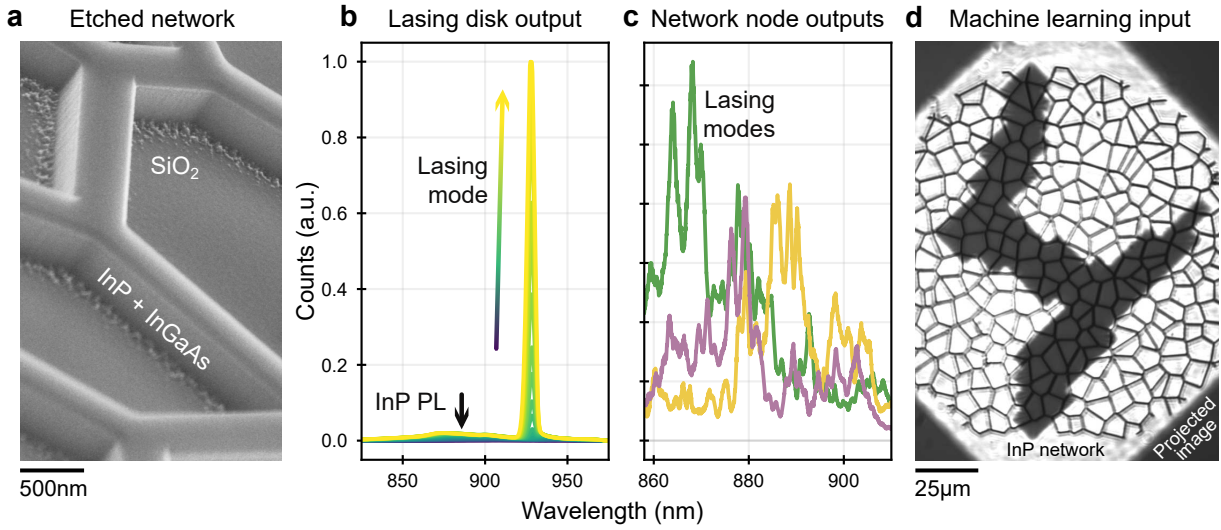
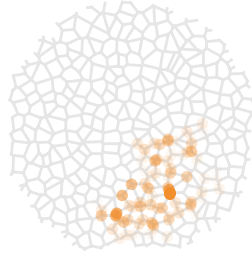


FIGURE 1.7: Thesis outline. **a** Etched III-V semiconductor network. Waveguides of InP (in this case with embedded InGaAs *quantum wells* (QWs)) join at nodes, forming a lattice. III-V etching will be discussed in Chapter 2. **b** Lasing spectrum from an InP disk device. Semiconductor lasers and the use of lasing as a fabrication diagnostic is covered in Chapter 3. **c** Lasing modes outcoupling from three nodes of a semiconductor laser. The physics of InP network lasers are explored experimentally and theoretically in Chapter 4. **d** A handwritten 4 is projected onto a network laser. The resulting spectrum can be used to predict what the projected number was. The application of network lasers to machine learning is tested in Chapter 5.

validation. I show lasing results from microdisk nanolasers that support *Whispering Gallery Modes* (WGMs), which prove to be a stable platform that can be analysed statistically to gain insight about chip quality and whether fabrication steps improve device performance.

Chapter 4 explores the physics of semiconductor network lasers. I look at their spectra, the shapes of the lasing modes, and how the waveguide dimensions impact the transport of light and losses within the network. With the help of simulations, I explain mode competition and how it leads to highly non-linear behaviours.

Chapter 5 showcases the lasing network as a platform for neuromorphic computing. I start by showing feature detection, a task that closely matches the mechanics of mode competition within the laser. I show that this application gives us insights into the underlying physics of the laser, and demonstrate negative responses to input, uncommon in optical systems. We then move onto classification tasks, where logistic regression is used to ‘train’ a simple classifier on the laser’s output. Two architectures are demonstrated: a single layer approach, and a two layer photonic convolutional neural network, which uses feature detection to boost performance. The system performs well on tasks like digit and clothing recognition, and breast cancer diagnosis, especially when only limited training data is available. At the end, I explore the likely physical basis for our strong performance, including our system’s similarity to ganglion cells in the retina.



Chapter 2

III-V fabrication and optimisation

Parts of this chapter concerning InP plasma etching were adapted and expanded from the following publication, for which I performed the majority of fabrication and measurements, produced the figures, and wrote the manuscript:

J. Dranczewski, A. Fischer, P. Tiwari, M. Scherrer, D. Saxena, H. Schmid, R. Sapienza, K. Moselund, “Plasma etching for fabrication of complex nanophotonic lasers from bonded InP semiconductor layers” *Micro and Nano Engineering*, vol. 19, p. 100196, Jun. 2023. [CC BY 4.0]

Abstract

The first step of bringing network lasers to the semiconductor platform is to establish a nanofabrication workflow. *Direct Wafer Bonding* (DWB) is introduced first, which allows us to integrate III-V semiconductors on silicon wafers. Lithography and plasma etching are then used to shape the semiconductor devices. The physics and challenges of plasma etching III-V materials in a research cleanroom are covered, and then a number of etch chemistries for InP are explored. Cl_2 -based etching is standard for InP but challenging at low temperatures. CH_4 -based etching works better at 80°C , but can leave polymer deposits. After showing results from a cyclic CH_4 recipe, which cleans the polymer deposits effectively, but leaves horizontal marks on the device side-wall, a continuous CH_4 etch is studied. A *Design of Experiment* (DoE) study allows us to discern the impact of etch parameters on device outcomes, and this insight is applied to optimise the continuous recipe for etching taller InP structures with InGaAs *quantum well* (QW) layers embedded.

2.1 III-V semiconductors and devices

Silicon is the semiconductor most often used in electronics, but it has an indirect band gap. This means that it cannot easily emit or detect photons through electrons jumping between energy levels, and other materials have to be used for applications in active photonics. The main class of materials of interest for us is *III-V semiconductors* – a class of semiconductors combining atoms from groups III and V of the periodic table, which have direct bandgaps and can thus emit light at a broad range of wavelengths.^[68]

Nanolaser systems from III-V semiconductor materials can show efficient lasing^[69–73] and are promising as part of on-chip nanophotonic systems. Potential applications of these systems include controllable on-chip light sources, sensing,^[74] and physical layers for machine learning processes.^[52] For these applications, we would be interested in marrying the worlds of electronics and photonics by integrating III-Vs on a Si platform. This is unfortunately difficult due to lattice and thermal mismatch between the materials, preventing direct growth of III-V layers on Si.^[74, 75] Progress is being made though, and many approaches to overcoming this challenge have now been developed.

Bottom-up integration approaches involve the assembly of material on the nanoscale to produce larger devices.^[76] Examples of these techniques include aspect ratio trapping,^[77, 78] nanowire growth,^[79] and *Template-Assisted Selective Epitaxy (TASE)*.^[80] These methods have advantages: for example, directed *metalorganic chemical vapour deposition (MOCVD)* growth, where material is constructed layer-by-layer,^[81] results in single-crystalline structures with extremely smooth and vertical side-walls.^[80, 82] This is important for lasing devices as smoother side-walls decrease scattering losses, vertical walls allow for efficient evanescent coupling, and the single-crystal structure has no defects where carriers would be lost. However, these techniques are time-consuming and challenging for complex shapes like the networks, and the design reproducibility can be poor.

We focus instead on a top-down approach, in which the devices are defined using lithography on a uniform area of active material, and the material is removed from regions where it is not needed.^[76] This allows for simultaneous fabrication of multiple complex, computer-defined designs such as coupled laser arrays, hyperuniform networks, and random lasers. A high degree of control and reproducibility is another benefit – the devices will reliably follow the lithographically-defined shape. The top-down approach comes with its own challenges too: etching III-V material can be complex, narrow trenches demand vertical side-walls and decrease etch rates, and waveguides

need high quality sidewalls to limit losses as light propagates over large distances.

Due to the complex shapes of our devices, and our desire to fabricate chips that sweep a large variety of device parameters, we have chosen to focus on top-down fabrication for this project. In this chapter, I will explain the particular process we followed, as well as the steps taken to overcome the aforementioned fabrication challenges.

2.2 Direct Wafer Bonding for III-V integration

Our III-V integration method of choice is *Direct Wafer Bonding* (DWB), enabling top-down fabrication of patterned devices. In DWB an epitaxially-grown layer of III-V active material is bonded to a Si wafer, avoiding the lattice mismatch defects that would arise if a III-V layer was to be directly grown on Si. The process is illustrated in Fig. 2.1. The active III-V layer is epitaxially grown on a lattice-matched III-V substrate with an etch stop layer in between. This is then turned upside down and bonded to a prepared Si wafer through a buffer layer of SiO₂ (which is deposited on both wafers) and annealing.^[83–85] The *burried oxide* (BOX) SiO₂ layer also optically isolates the III-V from the Si, leading to good confinement in the active layer. The sacrificial III-V wafer is then removed by wet etching, leaving the active layer attached to the Si wafer. Lithography is used to create a mask corresponding to the desired devices, and the patterns can be transferred into the active medium using plasma etching. As the final step, the etched devices can be processed chemically and encapsulated to help minimise the effects of surface roughness and further oxidation.^[86]

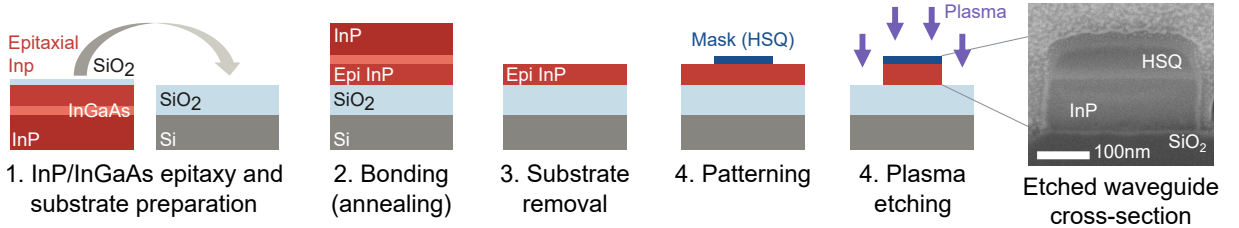


FIGURE 2.1: Illustration of the direct wafer bonding process described in the main text. A *Scanning Electron Microscope* (SEM) cross-section of a waveguide produced using these steps is shown on the right.

Here we have used InP active layers, but since the active layer is grown through MOCVD, more complex make-up is also possible, such as using a layer of InP with embedded InGaAs QW layers. While the main focus of this thesis will be the InP devices, I will also showcase some etching and lasing results from QW lasers to demonstrate the versatility of this process.

For patterning the mask that determines where material should remain after etching we have used optical and electron beam (e-beam) lithography. For etch test samples I initially used an optical lithography process, which uses fixed reflective masks that only let light through in the regions where we want to expose and cross-link the resist. This cross-linked pattern can then be transferred into an SiO_2 ‘hard mask’. The pattern used here was a series of small, repeating ‘cells’ that each contained long lines and other test patterns – see Fig. 2.2 a. These could then be cleaved (broken apart cleanly along a crystal plane) into separate test chips.

Later etch test chips, as well as the samples used for lasing tests, were patterned with e-beam lithography, where electric fields are used to shape a beam of electrons to match the designed shapes. This gives the higher resolution needed for some of the devices of interest, and allows for arbitrary and parametrically-varying computer-aided designs with quick iteration. An example of a digital mask design is shown in Fig. 2.2 b.

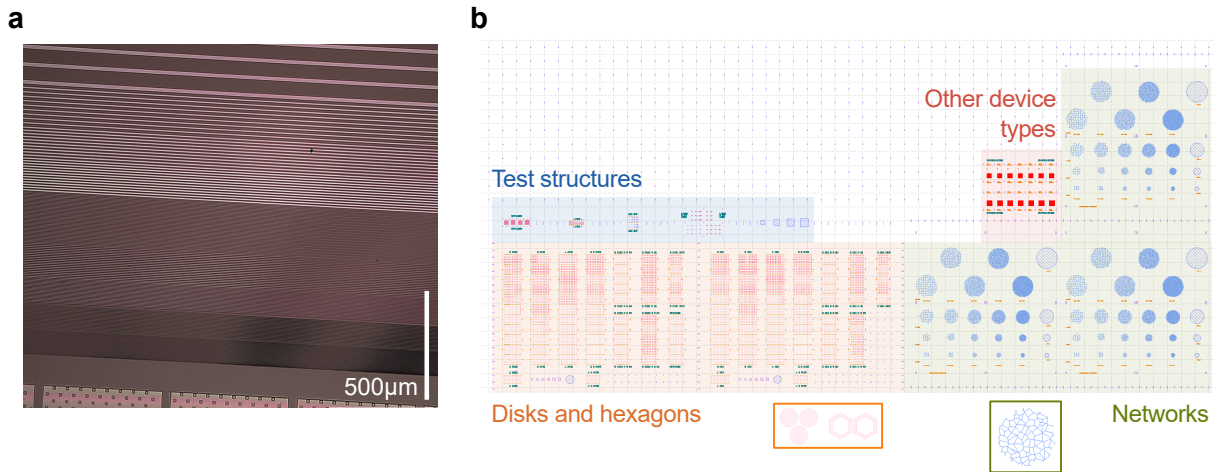


FIGURE 2.2: Lithography for mask creation. **a** Optical lithography with a fixed mask can be used to create repetitive designs. Here a test chip is shown, with a mask consisting of long horizontal lines that can be cleaved across post-etch to obtain a waveguide-like cross-section. **b** Computer-aided mask design for e-beam lithography allows for more flexibility, with parametric sweeps of disk lasers and lasing networks shown here.

Wafers or chips were coated with a 3 nm SiO_2 adhesion layer for the e-beam resist using *Atomic Layer Deposition* (ALD), and *hydrogen silsesquioxane* (HSQ) was deposited on top with spin-coating. The mask thickness used depended on the exact test and varied from 30 nm to 150 nm. This was then exposed with a Vistec EBPG 5200+ e-beam lithography machine, and developed. A short *Reactive Ion Etching* (RIE) etch removes the adhesion layer, and then the cross-linked HSQ can act as a hard mask directly.

Once a mask is patterned on a chip, the active material that is not covered by the mask needs

to be removed. We have opted for plasma etching for this, as it allows for vertical walls and a high degree of control over the etching process. This is in contrast to ‘wet’ chemical etching for example, where chemical reaction with a liquid are used to remove material, resulting in etched side walls that are often more organic in shape (either sloped or under-etched below the mask).^[87] The principles of plasma etching are discussed in detail in Section 2.3.

Finally, after the chip is etched, we need to ascertain what the results of the etching look like. The tools used for this are illustrated in Fig. 2.3 a. The long lines patterned with optical lithography can be inspected with a profilometer (in our case a Bruker DektakXT), which is a fine-point needle that detects changes in height on the order of nanometres. This gives an initial idea of the etch depth. For a more thorough observation, the chip can be cleaved, loaded into a *Scanning Electron Microscope* (SEM) vertically, and the cross-section can be observed directly, as seen in Fig. 2.3 b. This gives a detailed idea of the etch depth, wall angles, surface quality, and mask degradation.

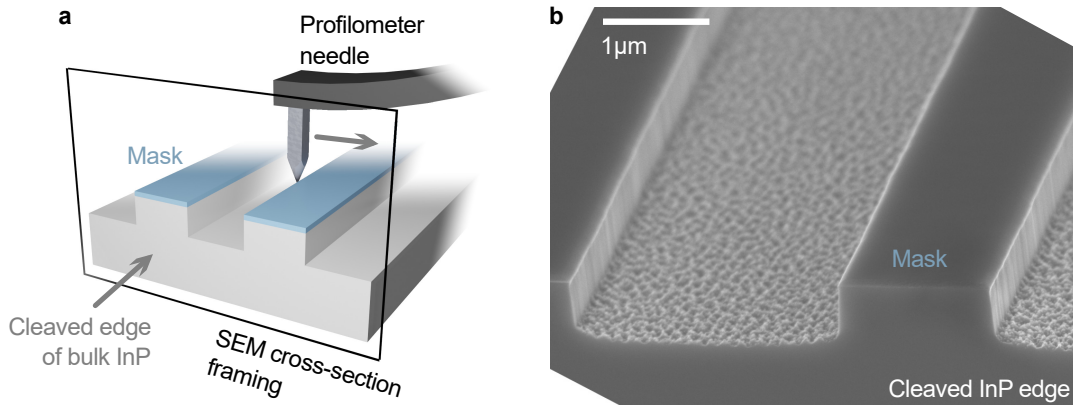


FIGURE 2.3: Tools used to diagnose the etching process. **a** With lines of varying width and separation masked out and etched on a chip, we can cleave it in half to observe the cross-section of the etch with a *Scanning Electron Microscope* (SEM). For a quicker inspection right after etching, a profilometer can be used – a needle (drawn not to scale) is dragged across the lines, and we can ascertain the etch depth based on its deflection. **b** SEM profile of an etched and cleaved InP chip.

2.3 Principles of plasma etching

In the general process of plasma etching, a variety of gases are mixed and excited into a plasma state and accelerated towards the surface of the sample. The removal of material happens through a combination of chemical and physical interactions of the etching plasma with the material’s surface.^[88] Chemical reactions of the active etching species with the surface results in formation of new molecules, some of which will be volatile and desorb back into the plasma

phase. Passive ion bombardment does not chemically react with the material, but it delivers energy to the surface, increasing the rate of chemical reactions and the evaporation of volatile compounds,^[88,89] as well as physically knocking material off. The inclusion of passive species travelling perpendicular to the sample also ensures that the etching is anisotropic, especially if the chemical etching produces non-volatile species – these will prevent further chemical etching of the device side-walls, but will be physically knocked away outside of the mask, allowing for etching downwards, but not sideways (this is referred to as *sidewall passivation*^[90]).

The machine used for plasma etching within this thesis was the Oxford Instruments Plasmalab 100, an *Inductively Coupled Plasma (ICP) Reactive Ion Etching (RIE)* tool, which is illustrated in Fig. 2.4. The plasma in an *ICP RIE* system is generated by a coil with alternating current running through it – this induces changing electromagnetic fields, which in turn induce currents in the chamber gasses, exciting them to a plasma.^[91] The plasma is then accelerated towards the sample by a separate *Radio Frequency (RF)* generator connected to the wafer holder. This allows for semi-independent control of the energy of the plasma (through the coil power, referred to as *ICP power*) and the speed at which it hits the sample (through the wafer holder generator’s *RF power*), providing a means for balancing chemical and physical etching.

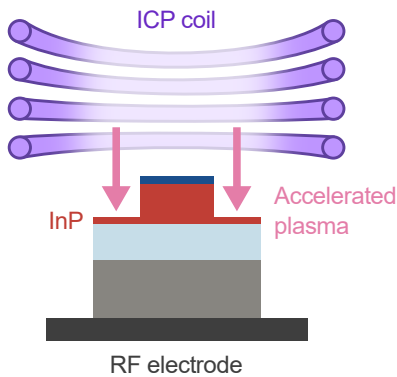


FIGURE 2.4: The chamber of an *ICP RIE* device, illustrated. The plasma is generated by the coil in the space above the sample, and then accelerated downwards by the RF electrode.

Another important parameter an *ICP* machine allows us to vary is the chemical composition of the plasma, and various potential chemistries for etching III-V semiconductors exist. To understand how these operate, let us begin by looking at the products that the plasma gases produce when interacting with the surface. Table 2.5 summarises the volatility of the products of In and P (as well as Ga and As) with different etching plasma species. The boiling temperature is not the only factor that is relevant, but it is a good indicator of how effective the recipe will be on a chemical level.

Element	Boiling temperature		
	Chlorides	Hydrides/trimethyls	Fluorides
In	InCl ₃ : 300 °C _(s)	In(CH ₃) ₃ : 55.7 °C	InF ₃ : >1200 °C
P	PCl ₃ : 75 °C	PH ₃ : -87.7 °C	PF ₃ : -101.5 °C
	PCl ₅ : 162 °C _(s)		PF ₅ : -75 °C
Ga	GaCl ₃ : 201.3 °C	Ga(CH ₃) ₃ : 134 °C	GaF ₃ : 1000 °C
As	AsCl ₃ : 130.2 °C	AsH ₃ : -55 °C	AsF ₃ : -63 °C
			AsF ₅ : -53 °C

TABLE 2.5: Boiling points for the products of elements that our III-V semiconductors consist of – our interest is in etching InP bulk, and InGaAs quantum wells. Temperatures marked with an (s) are specifically points of sublimation. Data adapted from Ref. [88].

A common etching agent for InP is Cl₂, as it can result in good etch rates, smooth walls, and does not form unwanted deposits.^[92,93] Due to the low volatility of InCl₃, these processes often rely on the temperature of the etching chamber to be high (>150 °C).^[92] For lower chamber temperatures, CH₄ and H₂ are commonly used as etching agents, which form volatile compounds with In and P respectively (with boiling temperatures comparatively close to room temperature). Unfortunately CH₄ tends to form polymer deposits that can interfere with the etching.^[94] Fluoride compounds are commonly used for etching Si,^[87,95] but they form highly non-volatile compounds with indium and gallium.

In tuning the exact molecular composition of the plasma, the chamber pressure, as well as the plasma energy and directionality (to balance the chemical and physical aspects of the etch), one can try to optimise the etch to obtain several desirable features. The result should be anisotropic, with highly vertical sidewalls, enabling coupled devices to exist close together and waveguides in network lasers to have well-defined rectangular shapes, especially where they join. The surface of the material should be smooth, with no imperfections that can lead to guiding or recombination losses. The etching should be *selective* between the mask and the etched material, meaning that the mask will not be damaged and areas covered by it will not be etched. These objectives are interlinked, for example low selectivity can lead to sloped device side-walls, as the mask degrades from outside-in, exposing the material gradually.

Having defined the objectives for what we consider to be a ‘good’ plasma etch, let us move on to the specific processes I have tested to obtain high quality lasing devices.

2.4 Exploring etch recipes

The etch recipes we concentrate on here can be split into three main categories:

- low temperature chlorine-based chemistries,
- cyclic 80 °C chemistries based on CH₄ and H₂,
- continuous 80 °C chemistries based on CH₄ and H₂.

First, low temperature chlorine-based chemistries. The **ICP RIE** tool I used was located in the Binnig and Rohrer Nanotechnology Center^[96] cleanroom at IBM Research Europe - Zürich. As the cleanroom is used for research purposes, and the tool is used for a variety of different processes, mostly at room temperature, the maximum chamber temperature possible to achieve was 80 °C. This limits the use of common Cl₂ recipes, since we have seen in Table 2.5 that products of In and Cl₂ are not very volatile and require high temperatures to efficiently desorb from the substrate. This class of recipes aims to overcome the temperature challenge with ion bombardment or plasma heating of the chip. My attempts to reproduce results from literature^[93,97,98] were generally not successful in this category, which is a challenge in the field of mid-temperature plasma etching^[99] due to differences between **RIE** machines and chip properties that may change material heating parameters. In the end, I have settled on utilising and exploring the parameters of two recipes previously developed at IBM,^[100] similar to Ref. [101]. These use Cl₂ and CH₄ as etching agents, and Ar or N₂ for ion bombardment.

The second and third class of recipes use CH₄ and H₂ as etching agents, which form volatile compounds with In and P respectively at the 80 °C temperature we can reach in our **ICP** chamber.^[88] The second recipe I study here was previously used for etching of microdisk lasers.^[80,100,102] It overcomes the issue of polymer deposits forming when working with CH₄ by etching the material in 2 repeating cycles: 1 minute of etching, then 16 seconds of cleaning the chip with O₂ plasma.

The third approach I develop here overcomes the CH₄ polymerisation by etching at higher powers to increase sputtering-based polymer removal,^[103] with a possible O₂ cleaning step afterwards to remove lingering polymer deposits. To explore the parameters of this approach and optimise the etching results, I have used a Plackett-Burman factorial *Design of Experiment* (**DoE**) method, which enables the study of many parameters using comparatively few experimental runs.^[104]

2.4.1 Chlorine-based etching

Attempts at etching with chlorine-based recipes produced mostly side-walls that were smooth, but not vertical. As shown in Fig. 2.6, even with some optimisation efforts starting from recipe I, recipe II remained fairly slow and physical, with a low selectivity and wall verticality. Recipe III, which uses argon for bombardment, is more aggressive and high-energy, and therefore faster, but micro-trenching (where a small dip is etched in the material next to the structure) results from its physicality. The higher selectivity is likely a result of the higher proportion of CH_4 , and the increased ICP power.

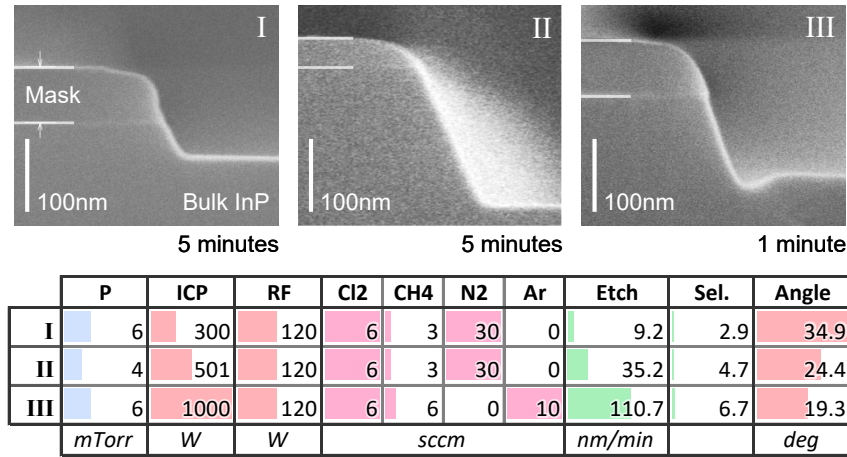


FIGURE 2.6: SEM cross-sections (see Fig. 2.3 for a reference illustration) and etch parameters for three ICP etching chemistries primarily based on Cl_2 (labelled with Roman numerals), etched at 20°C . The chamber pressure (marked P in the table) is maintained by the etching tool by tuning the venting rate to match the gas inflow rates (labelled here with the corresponding chemical formulas). A higher ‘strike’ pressure was used for the first seconds of the etch to help establish the plasma if needed. The *ICP* and *RF* powers correspond to the plasma generation and plasma acceleration powers respectively. The selectivity (*Sel.* in the table) compares the etch rate of InP with that of the SiO_2 mask, and *Angle* refers to the side-wall angle from vertical. The range of bar lengths in each column of the table is consistent in all figures throughout this chapter to enable comparisons between recipes.

Attempts at further variation, including different gas proportions, temperatures, and pressures resulted in similar or worse performance, with wall slant in the 20° to 25° range. While this is occasionally desired,^[105,106] the verticality of the walls is critical for our multi-waveguide devices, so the low volatility of InCl_3 in the temperatures we operate at appears to be a major obstacle.

2.4.2 Methane-based cyclic etching

Recipes using CH_4 and H_2 promise better performance thanks to the improved chemical desorption of the etched material at 80°C , at the cost of polymer deposits from the CH_4 . Cyclic

etching resolves this by alternating etching (1 minute) and cleaning (16 seconds) cycles until the desired etch depth is reached. The cleaning step cleans the surface from chemical deposits, and passivates the already etched side walls,^[90,107] which prevents chemical under-etching and results in highly vertical walls.

Fig. 2.7 presents a representative cross-section of a cyclic etch and recipe parameters. The cyclic and chemical nature of the etch is visible in the sidewall quality, as shown in Fig. 2.7 b and e. Horizontal lines are seen in SEM inspection of the test waveguides where the etching process switched between the etching and cleaning steps. Some general roughness is also visible, likely due to the etch relying largely on chemical desorption of material, which can be somewhat non-uniform – both ICP and RF powers are fairly low, so the plasma energy and vertical acceleration are relatively low too. The main disturbance is in this case along the length of a waveguide (see Fig. 2.7 e), and so should not impact its guiding ability too much. The roughness is small enough to not be seen under SEM inspection of both cleaved and *Focused Ion Beam* (FIB) waveguide cross-sections (Fig. 2.7 a and c respectively). Small deposits of likely condensed In have been seen on the walls in some samples as well, but treating the chip with a diluted H_3PO_4 bath removed these (see Fig. 2.7 b). Chemical post-etch treatments are discussed further in Section 3.4.3.

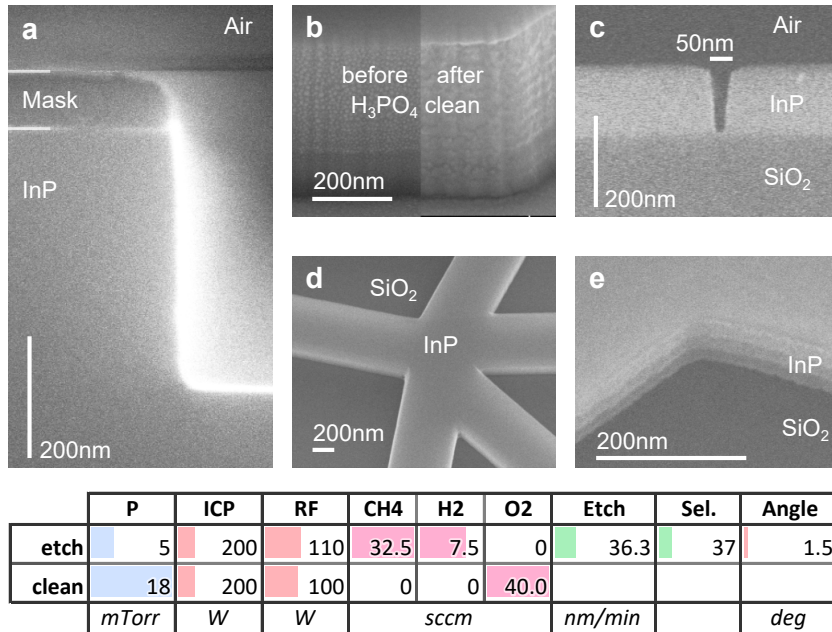


FIGURE 2.7: Etch results and parameters of the cyclic ICP etch recipe primarily based on CH_4 . **a** An SEM cross-section of a bulk InP etch, 12 cycles used. **b** Quality of the InP side-walls before and after a H_3PO_4 bath, with visible horizontal lines corresponding to etch cycles. **c** A FIB cross-section of a narrow trench etched in a layer of InP bonded on SiO_2 . **d** A junction between 5 waveguides, showing good transfer of the pattern into the active medium. **e** A closer view of the waveguide junction, with cyclic etch lines visible along the waveguides.

Combined with e-beam lithography, this process has proven to lead to very accurate representations of the devices designed in software. We observe a good reproduction of the network patterns on the micron scale (see Fig. 2.7 d), with waveguide widths and disk dimensions matching the design specification with a few nanometres of precision.

2.5 Etching Optimisation

In order to remove some of the artefacts of the cyclic etching, I was interested in investigating a continuous approach to etching InP with CH_4 and H_2 . This would remove the horizontal marks on the side walls, and also improve the overall etching speed – both thanks to higher plasma energies, and the ICP tool not having to switch to the cleaning chemistry every minute.

Using the same CH_4 etching chemistry as in the cyclic recipe, but running it continuously for a number of minutes, results in several issues (which can be seen in Fig. 2.9 I in Section 2.5.2, where I compare it to other etching results). The device wall angle is better than for Cl_2 but still worse than with cyclic etching. The horizontal roughness of the walls is largely removed, but the isotropic roughness (beyond vertical lines transferred from imperfections in the optical mask) remains, indicating that the issue arises from uneven chemical desorption. The etching is also very slow, and results in polymer deposits visibly settling on the chip (as a yellow tint), likely impeding etching and optical performance. Optimisation was needed to resolve these issues.

2.5.1 Design of Experiment principle

ICP etching is a process with many parameters that can be adjusted to potentially make the etch better. Unfortunately, using the tool is a time- and resource-consuming process, and so testing each of these parameters completely independently is not practical. *Design of Experiment* (DoE) is a field of study aiming to solve this precise problem – ascertaining the impact of many parameters on experiment outcomes by performing as few measurements (‘trials’) as possible.^[104,108,109] Approaches from that field are often used in etching literature to spot trends in how process parameters impact the results.^[90,93,98,110]

For this study, I have elected to use the Plackett-Burman method,^[111] which is illustrated in Fig. 2.8 a. This allows for performing the minimal number of measurements compared to the number of parameters we study,^[108] but this makes it suitable only for finding out the most significant effects in the system.^[109,111] It cannot ascertain the impact of interactions *between* the parameters for example, as is generally the case with fractional factorial DoE designs.^[104]

More measurements and other designs can be used to figure out these second-order effects (see for example Ref. [104] chapter 7).

For the study size chosen, 8 trials give insight into the impact of 7 parameters. Each parameter takes a ‘low’ value in half the trials, and a ‘high’ value in the other half. This way, the 4 ‘low’ and 4 ‘high’ trials can be averaged for each parameter to estimate the effect of going from ‘high’ to ‘low’ on various experimental outcomes, as illustrated in Fig. 2.8 b.

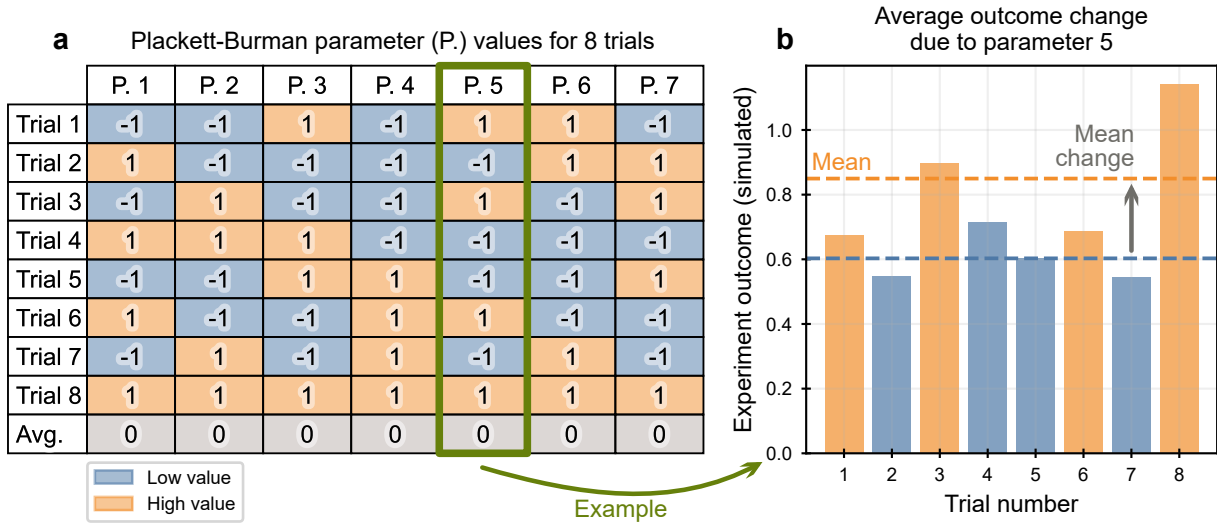


FIGURE 2.8: Plackett-Burman factorial *Design of Experiment* (DoE) principle. **a** The matrix showing values of the 7 parameters across 8 trials, generated with the Python package ‘pyDOE’.^[112] Each parameter has the same number of ‘low’ and ‘high’ trials. **b** A simulated example of analysing the impact of parameter 5 on the experiment. The ‘low’ and ‘high’ trials are averaged separately, giving an estimate for the average outcome change due to parameter 5 going from ‘low’ to ‘high’.

2.5.2 Methane-based continuous etching

I performed a Plackett-Burman factorial DoE study to roughly establish the influence of various etching parameters on the etch rate, selectivity, and wall angle. The parameters varied were the ICP and RF powers, chamber pressure, gas flow of CH₄, H₂, and Ar, and whether the sample was ‘clamped’ – glued to the carrier wafer, with He flow keeping the carrier thermally coupled to the ICP machine’s chuck and thus more steady at the set chamber temperature of 80 °C. 8 etch runs were performed on chips prepared with e-beam lithography, with parameter values determined by a PB matrix as seen in Section 2.5.1 above. The etched chips were cleaved, and etch quality assessed under SEM inspection. The resulting etch properties can then be selectively averaged to assess the mean impact of a property value changing from low to high.

The changes due the various parameters, and the definitions of ‘low’ and ‘high’ for each

parameter, are shown in Fig. 2.9, with illustrative SEM images of etches informed by this study shown in Fig. 2.10. The data gives some indication of the direction in which optimisation should go. We observe that higher ICP and RF powers lead to smooth side walls (compare Fig. 2.10 I and II), but going too high can be detrimental to the wall angle, and lead to uneven etching of the bulk InP, producing bulk surface roughness due to uneven In and P desorption.^[103] This roughness is less critical for bonded layers though, where etching for a little longer can remove the remnants of the material from the BOX surface. CH₄ is observed to play a key role in material removal, compared with the lesser impact of H₂ variation. A need for informed decisions in optimisation is illustrated as well – for example, we can opt for a lower etch rate if it leads to better wall angles.

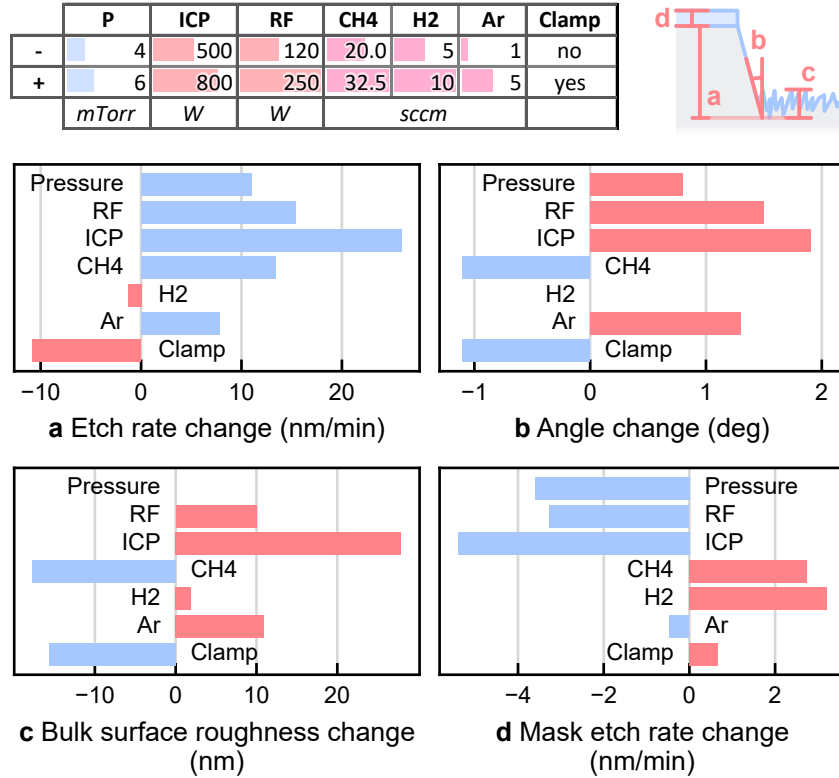


FIGURE 2.9: An overview of the Design of Experiment exploration of continuous CH₄-based etching. The first table shows the high and low values of the 7 parameters tested, and a-d show the average impact of going from low to high values of these parameters for four figures of merit (illustrated in the inset drawing). Blue bars indicate the desirable condition: lower angle, higher etch rate, etc.

Notably, the full impact of interactions *between* the parameters is not accounted for in a PB study with 8 experimental runs. Fig. 2.10 II-IV show one of the DoE etches (II), and two other attempts that explore other corners of the parameter space (III, IV). For example, it appears that lowering H₂ flow too much at higher pressures can lead to etches that are more isotropic

and produce small overhangs (Fig. 2.10 III). Lower pressures seem to work reasonably well with higher powers, producing one of the lower wall angles seen (Fig. 2.10 IV). This shows that after identifying the core parameters in the study above, it is possible to further tease out parameter relationships through experimentation or DoE methodologies varying fewer parameters.

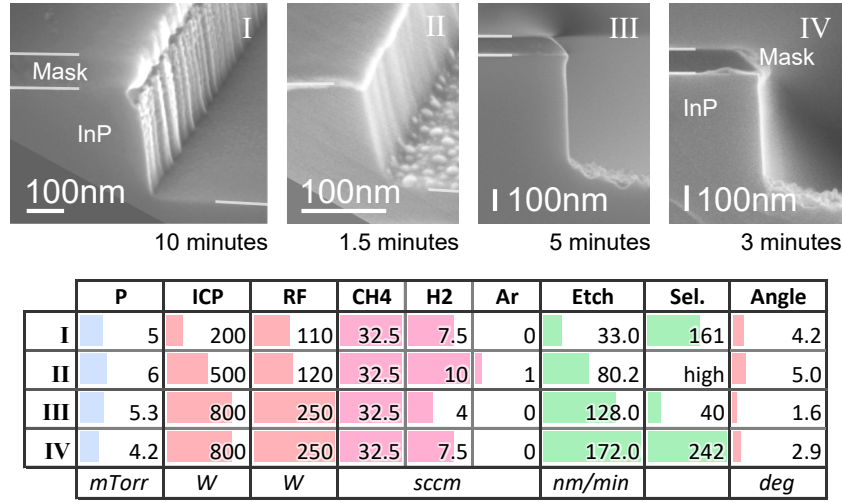


FIGURE 2.10: Etch results and parameters of four continuous ICP etch recipes primarily based on CH_4 . I illustrates the rough surface at low powers, II is one etch from the DoE study, III shows a slight overhang at high powers and low H_2 flow, and IV shows how lower pressures can lead to low wall angles.

2.6 Case study: InGaAs Quantum Wells

Since the active layer in the direct wafer bonding approach is epitaxially grown, we can alter the material composition of our lasing devices by introducing different steps in the growth process. For example, one can grow a layer of InP, and then several thin layers of InGaAs divided by InP spacing layers (see Fig 2.11 a and b for cross-sections). A stack like this acts as multiple one-dimensional *quantum wells* (QWs).

The larger band gap of the InGaAs wells results in a potential landscape within the material that encourages carriers to flow towards the wells and get confined in 2D, resulting in faster recombination and an altered emission wavelength (which can be further tuned through the size of the well layers).^[113] This can lead to more efficient devices, and ones that lase closer to the $1.3\mu\text{m}$ and $1.55\mu\text{m}$ wavelengths used in telecommunication.^[114] The use of quantum well-based gain material in lasers also leads to narrower lasing peaks in the output spectrum^[115] – I explore this more in Section 3.4.2. These advantages made the QW platform interesting to us, and it was a good case study to test the insights gained from InP etching and the DoE study.

The etching of InGaAs layers can use similar plasma chemistries, but there will be differences in etching performance due to material differences – see Fig 2.11 c for an example of how an InGaAs layer impacts the sidewall profile through the different etch rate, chemical attack rate, and wall angle, forming a discontinuity in the wall. Another challenge is that device scale has to be increased to match the higher lasing wavelengths, increasing etch depths and times. This can exacerbate differences in etch rates in narrow trenches or close to devices, making the uneven bulk surfaces produced by continuous CH_4 etching (originally seen in Fig. 2.10) more of a problem – see Fig. 2.11 d for an example where clumps of InP remain close to the waveguides.

With the insight into etching parameters from the DoE study, and a number of further etching tests, I was able to optimise the continuous recipe to work reliably for tall InP waveguides and disks that included InGaAs QWs. An example of a grating showing good etching quality is shown in Fig. 2.11 e. The QW layers for final samples are also fairly thin, so the cyclic CH_4 recipe managed to work with them without adjustments, and any wall non-uniformity that may have been the case was small compared to the horizontal features that arise due to the cyclic nature of the process.

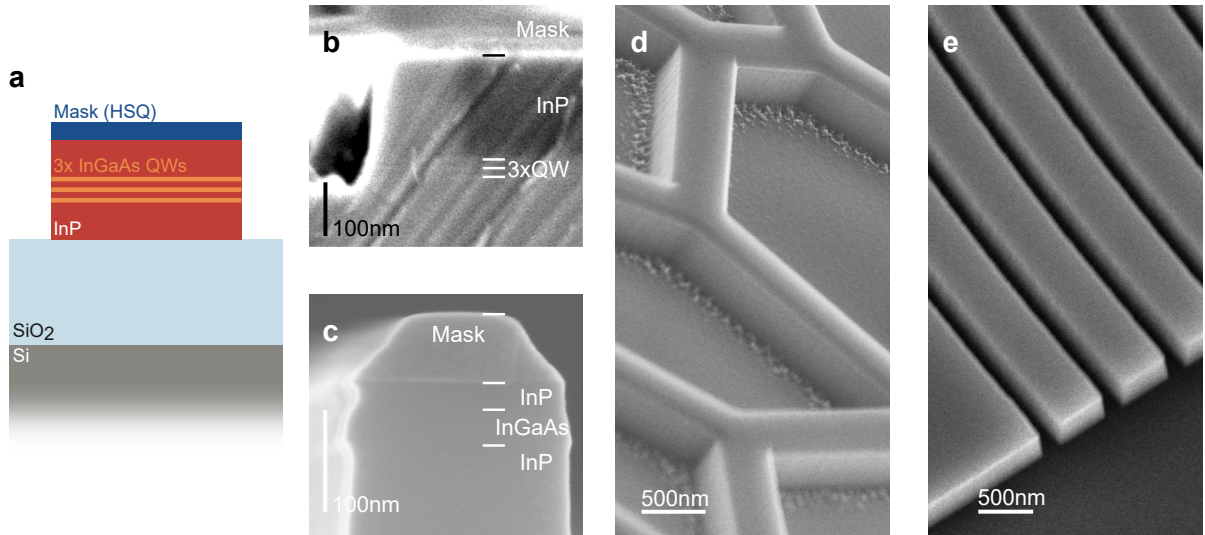


FIGURE 2.11: Quantum well etching. The material stack used for this process is illustrated in a and shown in SEM cross-section in b. The active material is wafer-bonded to a Si wafer through a SiO_2 layer, and consists of bulk InP with 3 InGaAs quantum wells embedded. c Different materials can respond differently to the same etching chemistry, making etching multi-material stacks challenging. Here, a thicker InGaAs layer visibly protrudes from the side of a waveguide (seen in cross-section). d An etch test showing smooth side walls, but problems removing residual InP around the base of the waveguides. e A final sample etched with a further optimised recipe, showing no residual InP, even within narrow trenches like in the grating coupler shown.

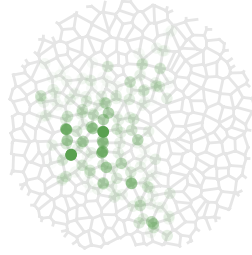
Conclusions

In this chapter, I have shown that we can perform high-quality III-V etching with the cleanroom tools available. Depending on the device and application, we can choose from continuous or cyclic etching, with Cl_2 or CH_4 chemistries. We can also transfer the etching process to new material stacks as required. With the DoE study in mind, and the tests performed on the QW material, the parameters of the best continuous CH_4 recipe I have used during my work are shown in Fig. 2.12.

	P	ICP	RF	CH4	H2	O2	Etch
etch	4	600	150	32.5	7	0	95.0
clean	18	200	110	0	0	40	
	<i>mTorr</i>	<i>W</i>	<i>W</i>	<i>sccm</i>			<i>nm/min</i>

FIGURE 2.12: The optimised continuous CH_4 etch chemistry. The ICP chamber is first conditioned by running the ‘etch’ step for 10 minutes. The sample is then inserted and etched. The etch time is based on the necessary etch depth – for bonded chips, I have added around 50 % more time to the etch step to make sure any remnant InP from the BOX surface and narrow gaps is removed, and account for differences from bulk InP. The cleaning step is run straight after the etch in the same chamber, for around a third of the etching time.

We will now move onto testing the lasing action from the many etched chips produced in this chapter.



Chapter 3

Semiconductor lasers for fabrication validation

Parts of this chapter concerning lasing from InP microdisks were adapted and expanded from the following publication, for which I performed the majority of fabrication and measurements, produced the figures, and wrote the manuscript:

J. Dranczewski, A. Fischer, P. Tiwari, M. Scherrer, D. Saxena, H. Schmid, R. Sapienza, K. Moselund, “Plasma etching for fabrication of complex nanophotonic lasers from bonded InP semiconductor layers” *Micro and Nano Engineering*, vol. 19, p. 100196, Jun. 2023. [CC BY 4.0]

Abstract

The fabrication optimisation in Chapter 2 focused on visible properties of the etched surface – wall angle and roughness, etch depth, etc. Since our desired application is a semiconductor laser, it is equally important to evaluate the devices’ lasing performance. This chapter introduces the principles of lasing, as well as how these apply to semiconductor nanolasers. We use microdisks as a basic, easily reproducible test platform, and show that InP and QW etched devices do lase. By fabricating arrays of disks with varying radii, we can sweep the InP gain spectrum, and use the lasing threshold as a proxy for cavity quality. Using this metric, we investigate the efficacy of O₂ plasma cleaning, H₃PO₄ treatment, and an Al₂O₃ passivation layer. We explore other possible lasing-based metrics, and showcase *continuous wave* (CW) lasing from the QW microdisks.

3.1 Introduction to lasing

Light-matter interactions can happen in a variety of ways, but the part most relevant to our application is how matter can absorb and emit photons. Fig. 3.1 a shows the three fundamental processes through which this happens.^[116] In the example here we will consider a simplified picture with two energy levels. In general, these processes can happen in systems with arbitrarily many energy levels – single atoms, molecules, or crystals, but the principle remains similar. When a photon is absorbed, an electron gets excited to a higher energy level, with the jump in energy corresponding to the energy of the photon. What is left at the lower energy level is a ‘hole’ – a lack of an electron. Even though this is a virtual particle, it is useful to consider, as it can be mobile in crystalline materials, and certain semiconductors can have more holes than free electrons to fill them.^[117] An electron can recombine with a hole spontaneously, emitting a photon in a random direction, and at a wavelength corresponding to the energy gap between the levels. This is called ‘spontaneous emission’. A process of more interest to us is ‘stimulated emission’. Here, a photon passes by an excited electron, prompting it to recombine with a hole and emit another photon. The new photon is coherent with the incoming one, which means that it is emitted in the same direction and with the same phase, or more generally into the same mode as the original photon.^[116,118]

If enough electrons in the gain medium are in the excited state, the system reaches ‘population inversion’ – a state where a photon travelling through the medium is more likely to encounter an excited photon than one in the low energy state. That photon is thus more likely to induce another emission rather than be absorbed. In this way, stimulated emission can be used to amplify the light travelling through a material. Population inversion can be achieved by injecting energy into the system, for example by external illumination. There are some requirements for sustaining population inversion, for example the system needs to have at least one other higher energy level that decays quickly to the second level. A detailed discussion of these conditions can be found in Ref. [116].

We can utilise the principles of stimulated emission to construct a ‘laser’, which stands for Light Amplification by Stimulated Emission of Radiation. Fundamentally, a laser is a cavity that traps light, such that it can be amplified through interactions with a gain medium, as shown in Fig. 3.1 b. The cavity supports multiple modes, and when their light goes through the gain medium, additional photons are emitted coherently through stimulated emission, and thus into the modes. Any light emitted not into a cavity mode (for example in a random direction through

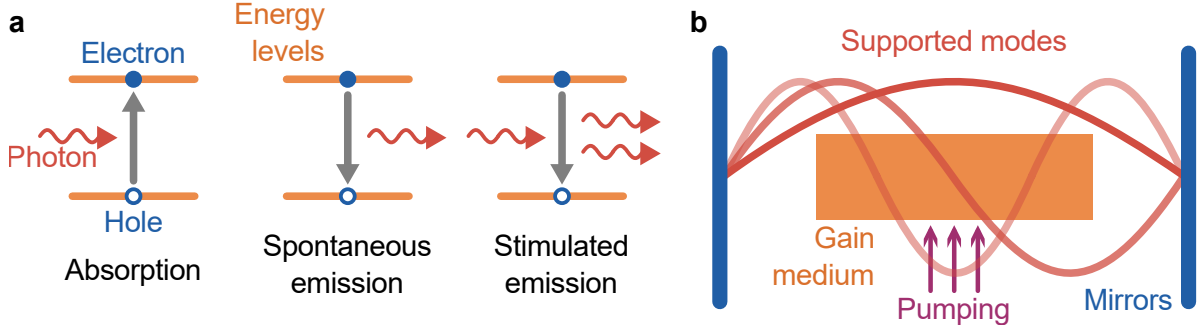


FIGURE 3.1: The core principles of lasing. **a** Kinds of basic light-matter interaction. **b** A simple laser is constructed out of two mirrors and a gain medium. The mirrors provide *confinement* and establish *modes* that the cavity can support (through phase-matching conditions). The gain medium provides *amplification* of the light within the cavity. We can ensure that the medium is in a state of population inversion by providing energy to it via *pumping*. Light that travels within one of the cavity modes (perpendicular to the mirrors in this toy example) will be continuously reflected and amplified, providing a strong and coherent beam. This is usually outcoupled by making one of the mirrors semi-transparent.

spontaneous emission) does not return to the gain medium, and is thus not amplified. Thanks to this selective amplification, the population-inverted laser emits light primarily into the cavity’s modes, creating coherent, directed light, like in the familiar laser pointer. If enough energy is pumped into the gain medium to overcome cavity round-trip losses, the system can reach a steady state and emit lasing light continuously – we call that energy the *lasing threshold*.

3.2 InP microdisk lasers

We have identified *confinement* and *amplification* as the primary factors that make a laser work. These are the principles that enable the archetypical two-mirror laser construction to function, but throughout the thesis I will move further and further away from this basic system to lasers that still confine and amplify light, but in ways quite different from a laser pointer. In fact, many nanolasers use methods other than simple reflection for confinement, from photonic crystals^[119, 120] to bound states in continuum.^[121] We will start our dive into nanolasers with whispering gallery mode microdisk lasers.

In a microdisk laser the semiconductor structure provides *both* the confinement and amplification. This principle is illustrated in Fig. 3.2. A wave travelling along the edge of a high refractive index microdisk will be reflected from its edges by *Total Internal Reflection (TIR)*. This confines the light to travel along the edge (see Fig. 3.2 a) and establishes *Whispering Gallery Modes (WGMs)*^[122] in the disk. These are named after the Whispering Gallery in St. Paul’s

Cathedral, which is a circular balcony below its dome. Thanks to the internal reflections from the gallery’s walls, speakers on one side of it can be heard on the opposite side,^[123] similar to how light is guided in the microdisk. Fig. 3.2 **b** shows how a microdisk laser looks like as a DWB etched device, with a simulated WGM overlaid. The confinement in a microdisk is not perfect, and some of the guided light scatters out of the device – this is the lasing output we will be measuring.

For the amplification side, a semiconductor will not have specific energy levels like a single atom illustrated in Fig. 3.1. Instead, interactions between electrons and the periodic potential of the crystalline lattice that they travel across result in the energy levels expanding into energy *bands*, where the electron’s energy depends on its wave vector. This band structure is illustrated for InP in Fig. 3.2 **c**. A notable feature typical for semiconductors is the ‘band gap’ – a range of energies where no electron states exist, indicated with a grey band in Fig. 3.2 **c**. III-V semiconductors specifically have a ‘direct’ band gap, which means that the closest bands below and above the gap approach each other at the same wave vector. An electron can thus jump

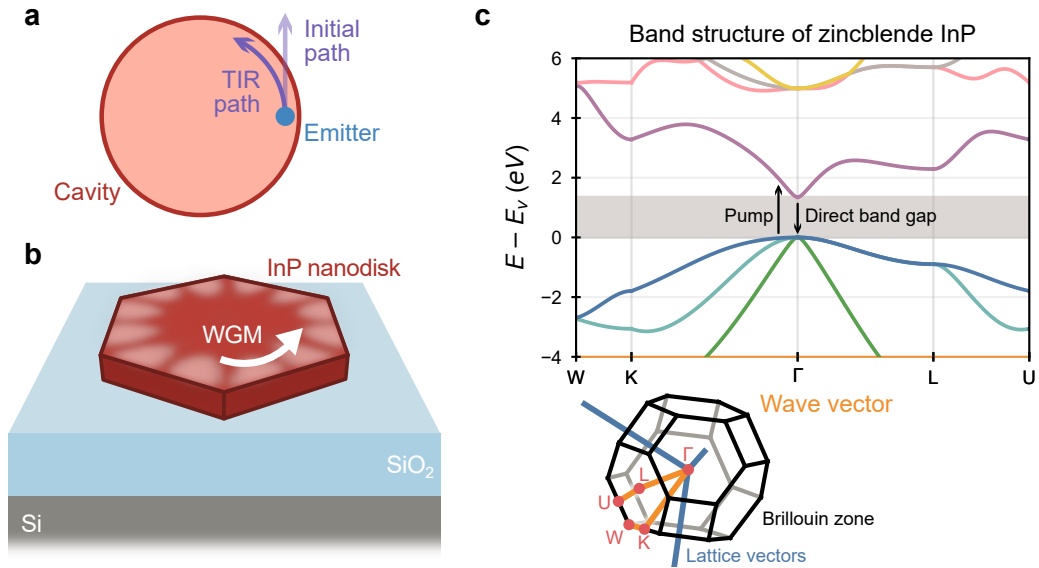


FIGURE 3.2: Confinement and amplification in an InP microdisk laser. **a** *Total Internal Reflection* (TIR) in a circular cavity results in a wave being guided along the edge of the cavity. **b** TIR guiding establishes *Whispering Gallery Modes* (WGMs) in a DWB microdisk. The mode profile is simulated with a *Finite-Difference Time-Domain* (FDTD) solver.^[124] The illustrated cavity is hexagonal to match bottom-up devices grown in IBM Research Europe – Zürich,^[80, 82] but circular cavities can be patterned and etched just as easily. **c** Simulated band structure for the face-centred cubic (‘zincblende’) crystal phase of InP. Energy displayed relative to the top of the valence band (E_v). Direct band gap and example pump light transitions indicated with arrows. Below is an illustrated key for the wave vector labels of the graph, showing the wave vectors of interest in a 3D representation of the crystal’s Brillouin zone.^[117] Data adapted from Ref. [125, 126] using Ref. [127]. A scissor shift^[128] was applied to the simulated band structure to match the experimental zincblende InP bandgap of ~ 1.35 eV.^[129]

from the valence (lower) to the conductive (upper) band (and vice versa) without changing its momentum. As only energy changes, this can be a purely optical transition, enabling direct band gap semiconductors to be used for light emission.

To use a III-V semiconductor for lasing, we need to achieve population inversion within it. In laser diodes this is achieved by electrical pumping, but in this thesis I will study optical pumping instead, as it allows for arbitrary device geometries and does not require specifically doped regions or metal contacts. To optically pump a semiconductor laser, we use light with photon energies above the band gap. This delivers electrons to the conduction band, and they can then thermalise down to the direct band gap crossing point, where the lasing emission happens.^[130,131] This is similar to a three-level approach to enabling population inversion,^[116] in that it allows us to deliver energy to the laser without de-exciting all of its electrons through emission stimulated by the pump photons – the pump light is at a different energy compared to the emission.

3.3 Experimental studies of lasing

Given that we will be optically pumping our nanolasers, we need an experimental setup to deliver pump light to the devices, and collect and spectrally analyse the light they emit. A schematic of the setup used at IBM Research Europe – Zürich is shown in Fig. 3.3. The pump is a ~ 30 ps pulsed supercontinuum laser with a repetition rate of 78 MHz. This means that the

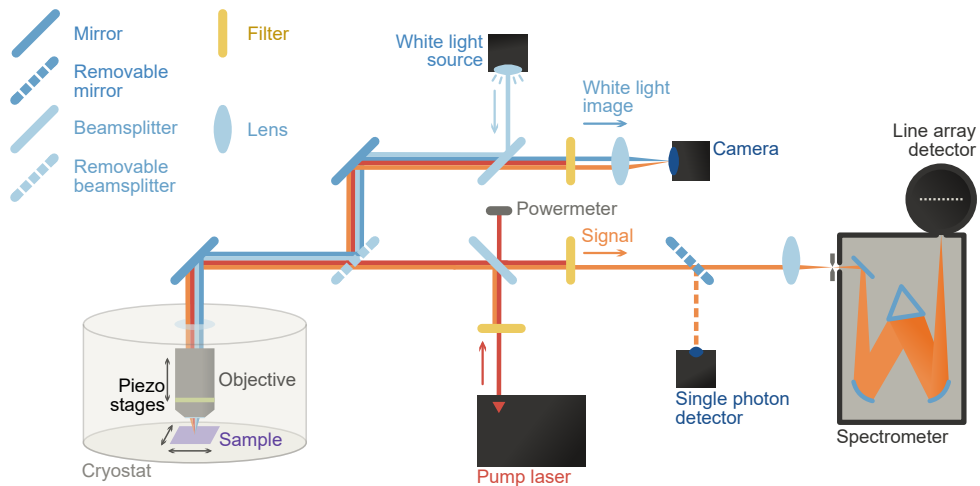


FIGURE 3.3: Diagram of the optical setup used for this investigation. The white light path is used for imaging and alignment and can be removed for measurement. A *Single Photon Detector* (SPD) can be inserted into the signal path instead of the spectrometer to measure the decay rates of device emission.

device generates a broad spectrum of lasing light, and specific parts of it can be chosen as the output. We can thus pump the InP devices at a wavelength of 750 nm (below the InP bandgap), but switch to ~ 1100 nm for QW devices incorporating InGaAs. The etched chips are positioned on a stage that uses stick-slip piezoelectric actuators,^[132] which have a large range of motion at sub-micrometer positioning precision. An objective is used to focus the pump light onto selected devices, and collect the emitted radiation at room temperature. The setup also allows for cooling down the chip in a cryostat down to 100 K or below using liquid nitrogen.

The spot size of our pump laser is ~ 1 μm in diameter, which is comparable or smaller than all the devices analysed, so we assume all of the pulse power is delivered to the device and use pJ /pulse as the pump energy unit. For more details on the pumping efficiency, Appendix Fig. B.1 shows a small study of how the device threshold changes with the pump location.

As we gradually change the power of the pumping laser, it will eventually be enough to sustain a population inversion in the InP cavity. Until then, the light emitted from the cavity will be normal, broad *photoluminescence* (PL) from spontaneous emission around the material's band gap. After the pump power is high enough to achieve population inversion, a narrow lasing peak arises from the PL, as seen in Fig. 3.4 a. Looking more closely at this peak (Fig. 3.4 b), we can see that a broader resonant mode is visible before the threshold as well, and then narrows significantly as we reach lasing threshold. We can fit the bulk PL around this peak (dashed line in Fig. 3.4 b) and subtract it (Fig. 3.4 c). We then plot the total integrated spectrometer counts under the lasing peak as a *light-in light-out* (LL) curve as seen in Fig. 3.4 d. The change in slope is characteristic of lasing, and by fitting a linear function to it we can extract a value for the lasing threshold. As the mode starts to lase, the PL of the bulk material is 'clamped' (does not rise any more), stimulated emission becoming more likely than spontaneous.

The inset in Fig. 3.4 d shows the far field emission of a disk device above the threshold, demonstrating fringes (partially highlighted orange) characteristic of coherent emission. The emission narrowing, intensity onset, gain clamping, and fringes in the far field emission are all signs that the device is lasing.^[55, 102, 133, 134]

The lasing peak in the spectrum in Fig. 3.4 a is broader than one would expect from a single-mode laser, and can be seen shifting towards lower wavelengths as the pump intensity is increased. These two phenomena can be explained by changes in refractive index of the material. The index can go up due to pump-induced heating,^[135] or go down due to a number of carrier-induced effects in InP.^[136] We find that the latter dominates in InP, resulting in the peak *blue-shifting* as the refractive index of the cavity decreases. As the pump laser is pulsed, this

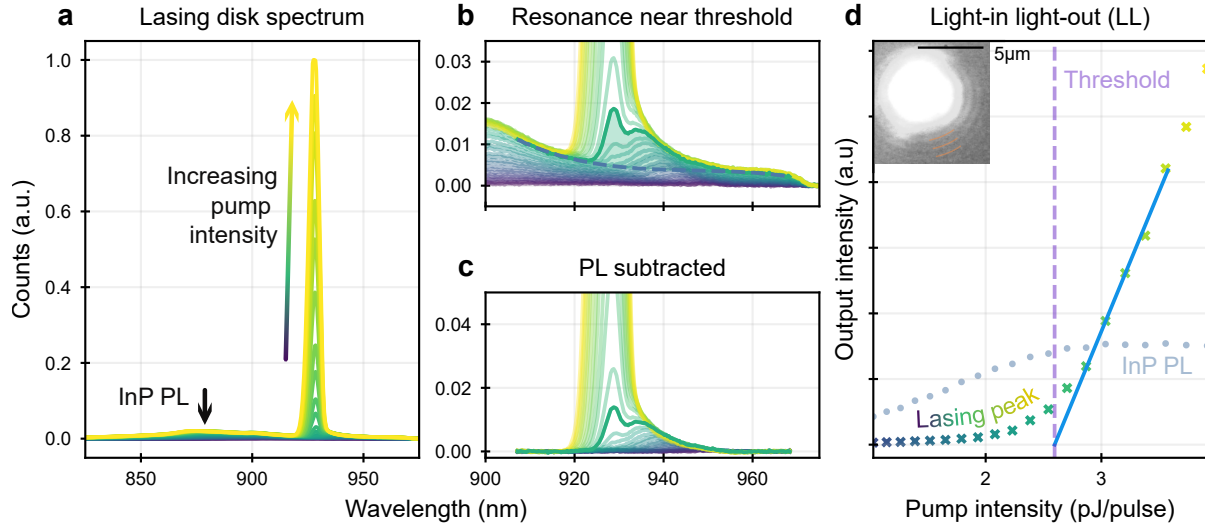


FIGURE 3.4: **a** A spectrum of a lasing peak emerging from the bulk PL of a hexagonal InP disk ($0.7\,\mu\text{m}$ in radius) as the pump laser intensity increases. **b** One spectrum is highlighted in green – there, a narrower peak emerges from a cavity resonance on top of the bulk PL (dashed fit line). As the power rises, this peak dominates the bulk PL. **c** The fitted PL can be subtracted to obtain the intensity of the resonance and the lasing peak alone. **d** Integrating the signal remaining after PL subtraction, we can obtain the intensity of resonant emission and lasing as a function of pump power (plotted as crosses matching the line colours in *a*). Fitting above the inflection point lets us extract the lasing threshold, with the far field emission of a disk device above this threshold shown in the inset. The bulk PL, shown as grey dots, is clamped once the device starts lasing.

blue-shift happens multiple times over a single spectrometer exposure, resulting in wavelength chirping and the peak appearing broader than it would be in steady-state.^[137]

3.4 Lasing threshold statistics for fabrication evaluation

In Chapter 2 I have demonstrated how *Scanning Electron Microscope* (SEM) inspection can be used to evaluate and compare the quality of plasma etching, including with quantitative metrics like wall angle and etch rate. Because our ultimate application for the etched devices is lasing, I was interested in exploring whether lasing performance by itself can be used as a metric to compare etching methods.

To define ‘lasing performance’ for the purposes of this investigation, I will be using the lasing threshold of microdisk cavities. Microdisks are a simple, stable platform, which can be easily added as a test pattern to any DWB etched chip. They can be laid out in a basic grid, and (in contrast to more complex nanophotonic devices) are quite easy to illuminate and align to – this allows for automation, like with the piezoelectric actuators in our setup, where the sample can be automatically moved and aligned to measure dozens of devices, giving statistical power to

the investigation. Sweeping many devices is also necessary as the location of the *Whispering Gallery Mode* (WGM) lasing peak in the gain spectrum will depend on the size of the cavity, and peaks closer to the centre of the gain curve will lase more eagerly. By fabricating and measuring devices with a range of diameters, we can sweep the gain spectrum and get a more complete picture. This idea is illustrated in Fig. 3.5. The location of the lasing peak for the microdisks goes up linearly with their diameter, as expected from phase-matching conditions for WGMs. A disk will support a comb of higher-order WGMs at different wavelengths, but only the modes close to the bandgap photon energy will lase. When the disk radius grows such that the mode of order N moves too far away from the bandgap wavelength, the $N + 1$ order mode starts lasing at a lower wavelength.

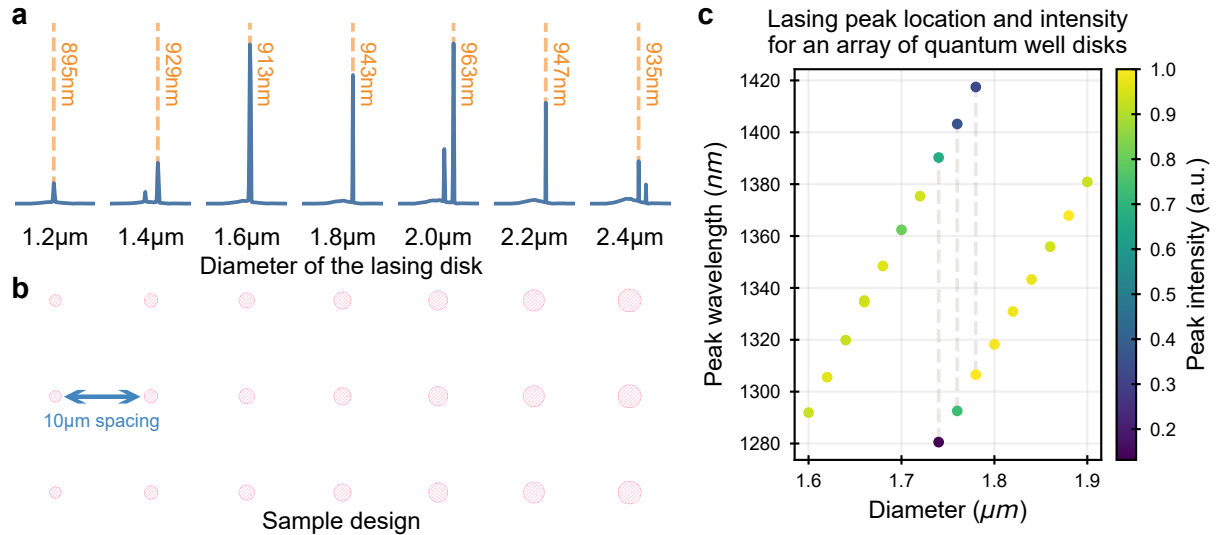


FIGURE 3.5: It is important to study disks with a range of emission wavelengths to cover the whole gain spectrum. **a** Lasing spectra for disks with increasing diameters, with the peak wavelength highlighted. **b** The device layout corresponding to the spectra above each column – digital design methods allow us to generate large parameter sweeps, which can then be etched onto chips and measured. **c** The lasing emission wavelength varies linearly with device diameter, as expected for a WGM laser. Dashed lines join peaks originating from the same disk.

The lasing threshold in a device is reached when the amplification of light through stimulated emission exactly matches the cavity losses. These can arise through imperfect trapping of the WGM, scattering on side wall imperfections, and non-radiative recombination of free carriers (Shockley–Read–Hall recombination in particular can happen through energy states created by lattice defects^[138]). We can thus say that the threshold is a good proxy for the overall cavity quality, and should correlate with how well lasing devices will perform on a given chip. The threshold does also depend on the emission wavelength and the gain volume available, but overall trends between chips can still be understood by measuring large numbers of devices that cover

these parameters.

3.4.1 Lasing from InP microdisks for different etching chemistries

To demonstrate the use of lasing threshold statistics as a fabrication diagnostics tool I will look at four samples here, fabricated with the cyclic CH_4 recipe, the variant of continuous CH_4 recipe shown in Fig. 2.12, and variants I and III of the Cl_2 recipes from Fig. 2.6. For microdisk devices to work, they need to be fabricated on DWB layers of InP, in contrast to the bulk InP used for etch tests in Chapter 2. The buried oxide layer used for these tests was $2\text{ }\mu\text{m}$. Compared to etching bulk material, the same recipes seem to etch bonded InP slightly slower, and with wall angles slightly larger, but the difference is not major, so we employ the same chemistry with etch times increased to compensate, and to make sure small trenches are resolved well.

Fig. 3.6 shows a comparison of the lasing thresholds extracted out of circular disk lasers on the four samples. Fig. 3.6 a shows the performance of the cyclic CH_4 etch, with most devices at thresholds in the 0.72 pJ/pulse to 3.0 pJ/pulse range, comparable or lower than those seen from etched and TASE-grown InP hexagons in a previous study, which were reported in the 1.81 pJ/pulse to 1.93 pJ/pulse range.^[80]

The continuous CH_4 etch devices shown in Fig. 3.6 b initially perform worse, but a further O_2 plasma clean brings the sample nearer to the baseline of the cyclic devices. This highlights both

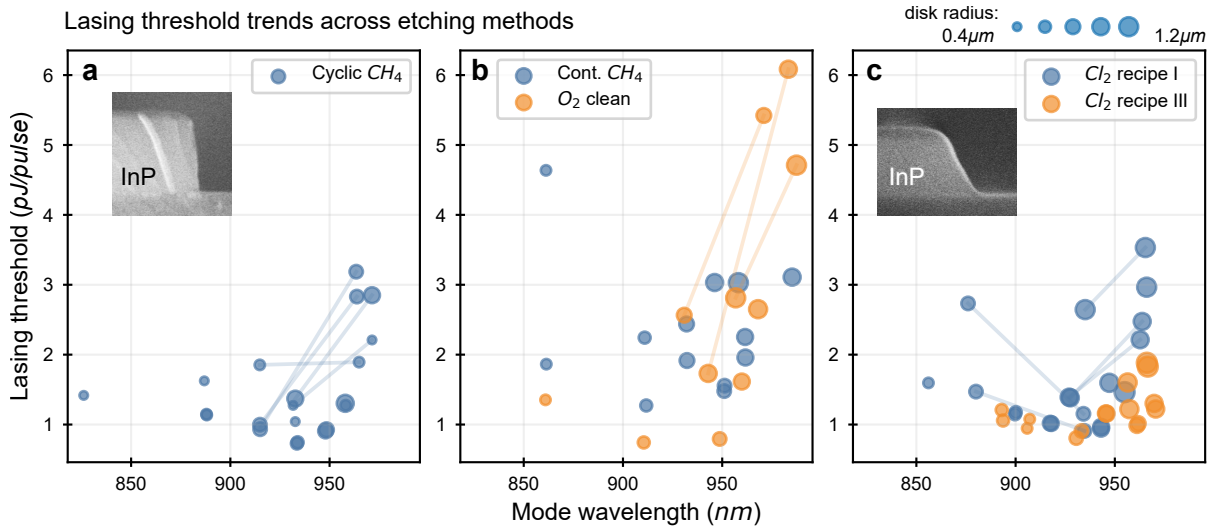


FIGURE 3.6: Comparison of the lasing threshold measured for circular disks on chips etched with different chemistries. The threshold is a function of wavelength (distance of the mode from the band gap), and a proxy for cavity quality. Some devices support two lasing modes within a single cavity – in that case the pair of modes is linked by a line on the plot. The sizes of the points correspond to the radius of the measured device. Inset SEM images show representative cross sections obtained with the recipes – highly vertical walls for CH_4 -based recipes, and smooth, sloped walls for Cl_2 .

that polymer deposits may not be entirely removed with higher power etching, and the value of lasing tests for highlighting fabrication issues that may not be visible under SEM inspection but are critical to our application.

Continuous Cl_2 etches are shown in Fig. 3.6 c, achieving performance comparable to cyclic etching. This shows that smooth cavities can lead to good lasing performance, despite the slanted walls impeding the total internal reflection necessary to support whispering gallery modes. The angles of the cavity walls still make the use of this chemistry difficult for coupled cavities, where narrow trenches are necessary, or for network designs, where many waveguides meet at a single point. It is worth noting that some applications do call for slanted-wall waveguides,^[105,106] so these recipes could be applicable there.

The use of cavity lasing as a proxy for fabrication fidelity is thus validated – here it highlighted the importance of cavity wall smoothness and established the need for post-etch O_2 cleaning for continuous CH_4 chemistries.

3.4.2 Quantum well lasing results

As mentioned in Chapter 2, QW lasers come with a number of advantages. Their emission wavelength is closer to those used in telecommunications, the carrier recombination within them is faster and more efficient, and devices can be more easily tuned to exact specifications by altering the well dimensions or the exact chemical make-up of ternary compounds.^[113,139] Most notably for lasers, the altered density of states in quantum well structures result in a lowered *linewidth enhancement factor*,^[115,140] which in part describes how much the refractive index of the material changes with the number of free carriers.^[141] Less change in refractive index means less blue-shift, and thus less chirp-induced broadening during pulsed operation. This can be clearly seen in the results from my etched microdisks in Fig. 3.7 a. In comparison to an InP peak, the one from a QW device starts significantly narrower, and does not broaden much as the pump intensity is increased. This feature is potentially advantageous for heavily multimodal systems (like our random lasing networks), as it makes it easier to distinguish closely-spaced peaks from each other.

Statistical lasing performance results for the 3x InGaAs QWs embedded in InP platform (shown in Fig. 3.7 b) look promising. The devices exhibit low thresholds of 1 pJ/pulse to 4 pJ/pulse, comparable to InP devices. The continuous CH_4 recipe I developed performs comparably or better with respect to the reference cyclic etch, especially for peaks located at lower wavelengths.

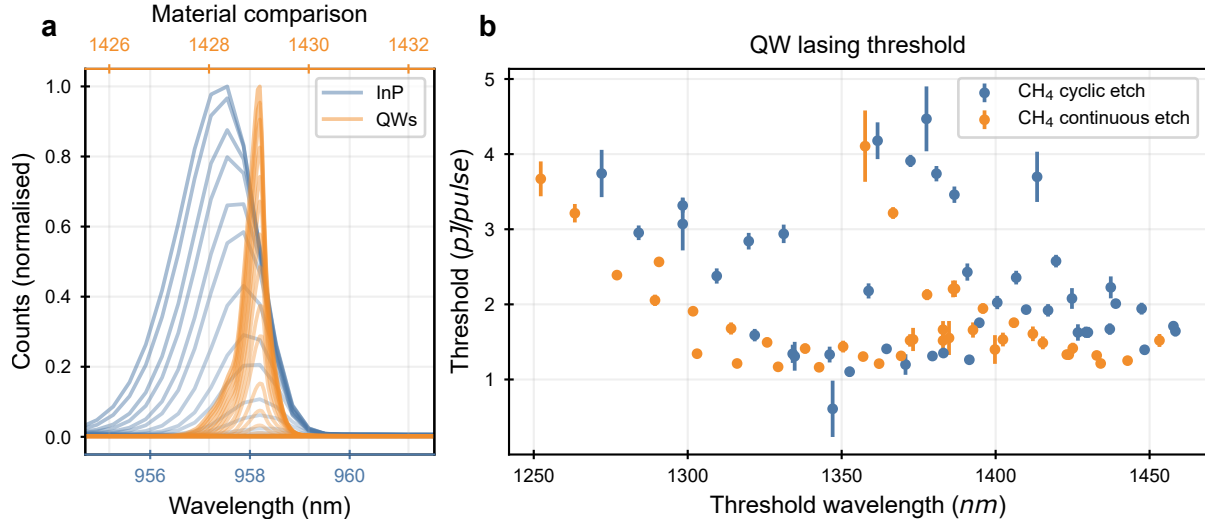


FIGURE 3.7: Quantum well lasing results. **a** A comparison of lasing peaks from etched InP and InP + QW microdisk lasers. The emission is at a different wavelength, and narrower for the sample with QWs. It also blue-shifts less as the pump power (line opacity) increases. **b** Lasing threshold statistics for an array of disks on two samples etched with CH₄-based chemistries.

3.4.3 Passivation and its impacts on cavity degradation

As we have seen, especially with the cyclic CH₄ side walls, plasma etching can produce surfaces that are physically and chemically rough, and leave various deposits. This can impede lasing performance, either through scattering, or non-radiative recombination losses. III-V devices are thus often chemically processed and passivated to improve performance after etching or growth,^[86] with the process claimed to reduce surface trap states and recombination velocity.^[142–144] This is especially important for InGaAs,^[86] but the exact degree of benefit for InP is disputed.^[86,144–147]

Many publications use sulphides for passivation, but ammonium sulphide is challenging to work with, and the exact preparation process varies and can be challenging to reproduce.^[143,144] We have thus opted for a simpler process where phosphoric acid (H₃PO₄, diluted with de-ionised water 1:10) is used to clean the surface of etching deposits (likely pooled In – cf. Fig. 2.7 b), and potentially passivate it^[148] and remove surface states.^[149] The devices were then fully enclosed in a 2.5 nm to 3.5 nm layer of aluminium oxide (Al₂O₃) through ALD, which further passivates them^[145,150,151] and prevents oxidation through encapsulation.

To test the impact of this procedure, I once again turned to lasing as a diagnostic. The experiment consisted of a 2.2 μm diameter, cyclic-etched QW lasing disk being continuously pumped with the pulsed laser for approximately 17 minutes. The spectral output was monitored throughout, and a LL curve taken approximately every minute to calculate the lasing threshold. The results are shown in Fig. 3.8. The H₃PO₄ bath lowers the threshold significantly, indicating

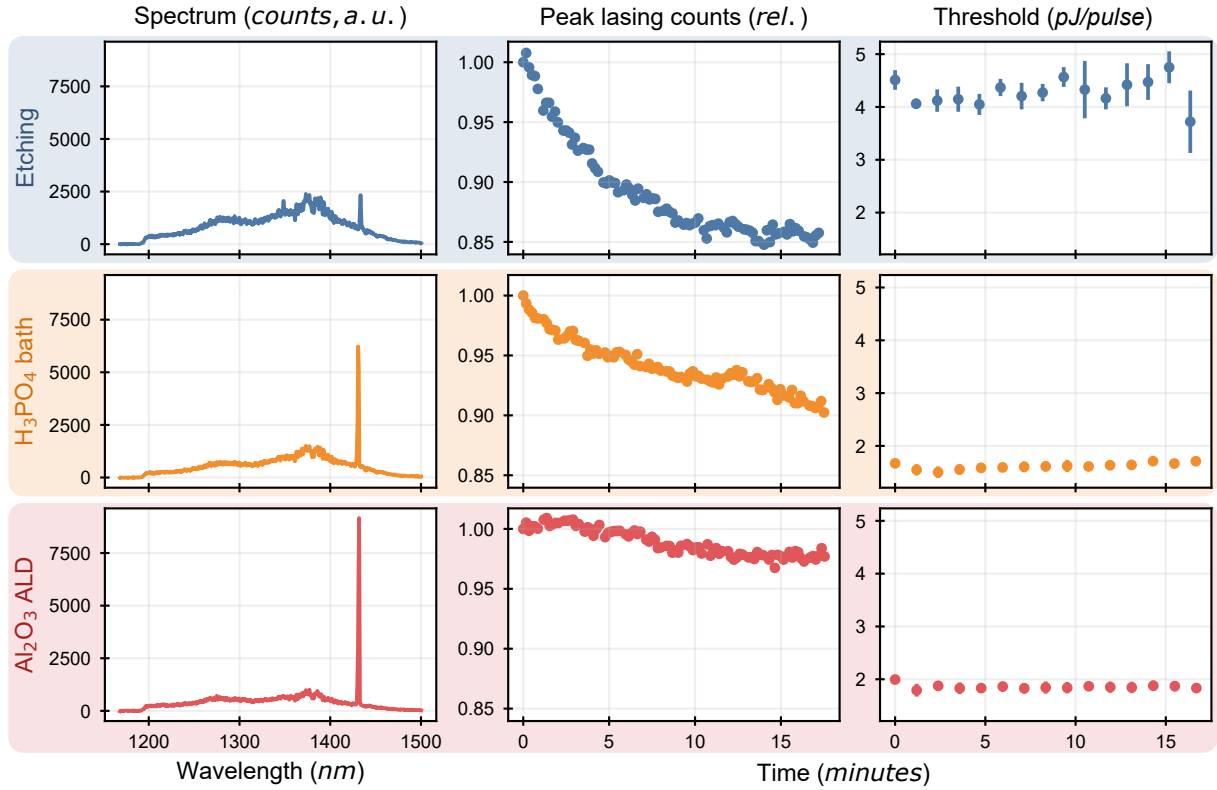


FIGURE 3.8: Measuring the impact of passivation on device degradation. The three columns show: the initial PL spectrum of a $2.2\mu\text{m}$ diameter, cyclic-etched QW lasing disk with a lasing peak visible, the height of the lasing peak over approximately 17 minutes of continuous pulsed pumping, and the lasing threshold of the peak, extracted by performing LL measurements at roughly 1 minute intervals during the experiment. Two treatments are applied (coloured rows), one after the other, with the same experiment performed after each treatment.

that the etching deposits significantly contribute to cavity losses. Both treatments generally improve the height of the lasing peak, and how much it decays under continuous illumination. The Al_2O_3 ALD coating may also stabilise the threshold value, though this is harder to ascertain.

This experiment validates the post-processing steps used by testing them in an applicable scenario. Degradation is a major concern for our devices, as using them for machine learning requires long periods of pumping as millions of unique data exposures are illuminated and measured. The good performance of the processed devices is a sign that they should work well for our application of interest, and the same testing procedure could be applied to trial other post-etch processing techniques to reduce degradation further.

3.5 Other lasing-based diagnostics

Throughout most of this chapter, I have focused on the lasing threshold under pulsed illumination as the main metric for lasing performance and fabrication success. In this final section I will briefly explore two other potential metrics that use lasing to illuminate different aspects of cavity quality: *continuous wave* (CW) lasing and emission lifetime.

Optically-pumped nanolasers are often studied under pulsed illumination, as the pumping heats the devices significantly.^[102,152] In pulsed operation the thermal energy can be conducted away between the pulses, but the heat is more difficult to conduct away in steady-state, CW operation, especially at room temperature.^[153] This can lead to rapid and permanent device damage.^[151]

Fig. 3.9 a shows a comparative study of QW microdisk CW lasing thresholds from chips obtained with cyclic and continuous CH₄ etching. The devices were pumped with a 1064 nm diode laser, and have thresholds of 150 μ W to 400 μ W. This is comparable to some grown microdisk devices, with InP-based platforms reaching 200 μ W thresholds,^[153] InAs *quantum dot* (QD) lasers working at 80 μ W to 300 μ W,^[154] and GaN/InGaAn QWs notably going down as low as ~ 7.9 μ W.^[155]

The good CW room-temperature performance of our etched disks compared to similar grown platforms is an indicator of low losses and good etch quality. We could not perform similar tests for the pure InP devices unfortunately, as our 1064 nm CW laser is above the InP bandgap.

Another lasing metric is the carrier lifetime. A cavity is excited with our ~ 30 ps pump pulse, and then the carriers within it decay over time, emitting light. Time-correlated single-photon counting can be employed to measure when these photons are emitted and detected with a *Single Photon Detector* (SPD) (see Fig. 3.3) and give an idea of how long it takes the cavity to de-excite.^[102,156] Fig. 3.9 b shows the main effect of interest – the lifetime of the QW microdisk emission shortens as we go to higher powers and it begins to lase. Its carriers are being de-excited more quickly by the stimulated emission.

While it is a good performance indicator that we can see the shortening at room temperature, this measurement is generally hard to scale and interpret further. Multiple factors like non-radiative recombination can influence the time-resolved behaviour, it is tricky to get enough signal to overcome the SPD noise floor, and the resolution is limited both by the length and shape of the pump pulse and the intrinsic time resolution of the setup (~ 50 ps). Nevertheless, the PL lifetime below threshold for the QW devices appears to be shorter than for similar InP

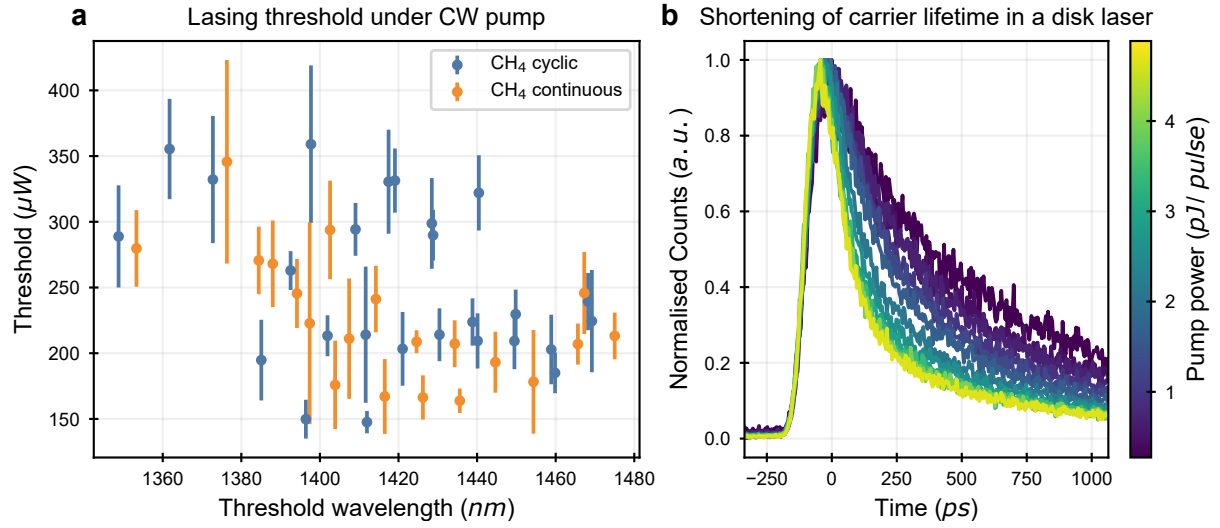
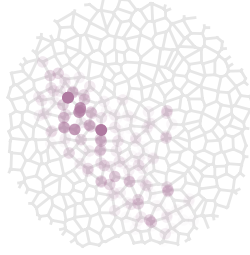


FIGURE 3.9: Other potential lasing-based fabrication diagnostics. **a** Lasing thresholds of disk devices on two chips under **CW** illumination. With no time to dissipate heat between pump laser pulses, this test can be a more direct measurement of the non-radiative recombination losses, and the thermal energy dissipation efficiency from the active medium. **b** Carrier lifetime measurements show how quickly the pulse-excited carriers decay in a laser. Stimulated emission results in the carriers decaying faster, which can be seen as the pump power is increased.

nanolasers,^[80] likely indicating efficient carrier recombination. The lifetime above threshold is similar, likely due to the **SPD** resolution limit.

Conclusions

After a successful fabrication study in Chapter 2, this chapter validates the optimised fabrication methods for lasing applications. Our disk devices lase well, and lasing threshold was demonstrated to be a good indicator of device performance, informing several fabrication decisions. The studies shown here have thus enabled me to optimise the etching process while keeping our end goal in mind – semiconductor network lasers.



Chapter 4

Semiconductor network lasers

Figures 4.1 and 4.11, parts of Fig. 4.5, as well as the related analysis, were adapted from the following publication, to which I contributed fabrication work, and some of the measurements:

D. Saxena,* A. Fischer,* J. Dranczewski, W. K. Ng, N. V. Trivino, H. Schmid, T. V. Raziman, A. Arnaudon, M. Barahona, O. Hess, K. Moselund, R. Sapienza, “[Designed Semiconductor Network Random Lasers](#),” *Laser & Photonics Reviews*, vol. 19, no. 1, p. 2400623, 2025. [CC BY 4.0]

Figures 4.6 and 4.9, as well as parts of the relevant analysis, are based on simulation and theoretical work by T. V. Raziman, and adapted from the following publication:

J. Dranczewski,* W. K. Ng,* A. Fischer,* T. V. Raziman, D. Saxena, T. Farchy, K. Stenning, J. Peters, H. Schmid, W. R. Branford, M. Barahona, K. Moselund, R. Sapienza, J. C. Gartside, “[Retinomorphic Machine Vision in a Network Laser](#),” *arXiv*, Aug. 2024. [*manuscript in review*]

*Authors contributed equally.

Abstract

Having established fabrication processes, and tested lasing from simple microdisk devices, we can now move on to studying the main system of interest in this thesis. This chapter demonstrates lasing from semiconductor network lasers and explores the physics behind it. We see that the spectra from these devices are highly complex, with many overlapping modes existing in different regions of the network. The conditions in which these modes arise is studied, looking at the losses at network nodes, and their dependence on waveguide width. We study the modes directly as well, observing their extent in simulation and experimentally. Facets of system non-linearity (relevant to the neuromorphic application) are studied, introducing the concept of mode competition to explain complex mode behaviours. Finally we test the sensitivity, stability, and flexibility of network lasers, bearing in mind their intended application.

4.1 System introduction

In Chapter 1, I have introduced the idea of network lasers – a random lasing platform where light travels through labyrinths of waveguides and is somewhat confined through multiple scattering at the nodes between them. As introduced in that chapter, the polymer-based platform the networks have been previously demonstrated on^[63,65] has disadvantages, like the lack of control over the device shape and bleaching of the dye used for gain. Creating a similar system on a semiconductor platform could help resolve these. Having explored III-V fabrication in Chapter 2 and the principles of semiconductor lasing in Chapter 3, we are now ready to realise the network laser systems in DWB InP.

Fig. 4.1 shows an overview of a realised device, with Fig. 4.1 a showing an SEM image of a fabricated network. The edges of the lithographically-patterned graph correspond to the boundaries of a Voronoi diagram’s regions for an assembly of semi-random points (where each region is the area where a specific seed point is the closest one). This generation method, as well as some potential alternatives, is discussed in more detail in Section 4.3.2.

Photonic modes in the network are formed by constructive interference across many paths^[63] as light is guided through the waveguides, and scatters between them many times at the network nodes (see Fig. 4.1 b). An example of a simulated mode is overlaid in Fig. 4.1 a.

Below the lasing threshold, a network will emit light uniformly from all the edges through spontaneous emission. Above threshold spontaneous emission takes over and most of the light is emitted into the network modes, and thus along the waveguides. A lasing network emission is thus mostly from the nodes where waveguides meet^[63] (see the right inset of Fig. 4.1 c and cf. Fig. 1.5 c), as that is where the light from the waveguides scatters out.

We can collect the scattered light and measure it through a spectrometer to obtain a lasing spectrum, as seen in Fig. 4.1 c. The one shown here is only from a section of the network, as the spectrometer input is cut with a slit to improve spectral resolution. Nevertheless, we can see that the spectrum is very complex, with many different lasing modes present above the normally broad *photoluminescence* (PL) of InP. These suffer from the spectral broadening characteristic of InP (as outlined in Section 3.3), with the effect of chirping likely even more pronounced than before – a 225 fs pump pulse is used here, in contrast to the ~ 30 ps one used for the microdisks (Appendix Fig. B.2 briefly explores the peaks narrowing when a longer pump pulse is used). This does make the specific lasing peaks more difficult to tell apart compared to the forest of sharp peaks characteristic of the polymer-and-dye-based networks.^[63] The left inset of Fig. 4.1 c

shows a clear ‘S’ shape of the LL curve, indicating a lasing onset at a threshold pump fluence of $60 \mu\text{J cm}^{-2} \text{ pulse}^{-1}$, further proving that the system is lasing.

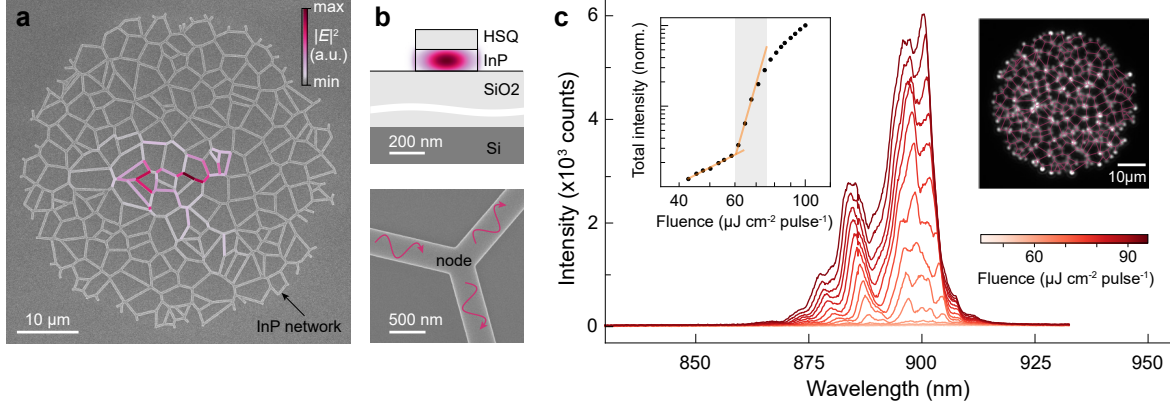


FIGURE 4.1: InP Semiconductor lasing networks overview. **a** SEM image of a fabricated InP network with $50 \mu\text{m}$ diameter and $2 \mu\text{m}$ average link length. An example of a simulated network mode profile is superimposed over the network image. **b** Each link of the network is a waveguide, as illustrated in the top panel. The light travels through the active material similar to an optical fibre – a simulated optical mode cross section is shown in the illustration. At the waveguide joints (SEM in the bottom panel), light scatters into adjacent waveguides. **c** Lasing spectrum from a section of the network (defined by a spectrometer slit) as the pump fluence is increased. Left inset shows the integrated intensity, showing a clear threshold characteristic of lasing. Right inset is a far field image, illustrating the guided light outcoupling at the graph nodes where waveguides join.

Fig. 4.2 shows an illustration of the experimental setup used to obtain the results above and in the rest of this chapter. It is similar in principle to the one shown in Fig. 3.3 but has additional capabilities for shaping the pump beam and for measuring the emitted light.

The pump laser is a 1030 nm 225 fs pulsed system with a repetition rate of up to 100 kHz . An *Optical Parametric Amplifier* (OPA) is used to convert this light to 633 nm (chosen to match the wavelength of a HeNe alignment laser), which is below the InP bandgap and can thus be used for pumping. The pulse energy is controlled by a variable *neutral density* (ND) filter mounted on a moving stage. Removable mirrors redirect the pump to an optional beam-shaping path, where the beam gets expanded, and hits a *Digital Micromirror Device* (DMD)^[157] – a mechanism consisting of 1280 by 800 microscopic square mirrors, which can be individually tilted between two orientations using electrostatic forces. This is the basis of the technology used in projectors, and allows us to discard selected parts of the beam to form black and white images out of it. The beam is expanded to a size greater than the DMD, so we can assume that the intensity across the device is approximately uniform. A series of lenses is used to project the resulting image from the DMD plane to the sample plane.

The sample is mounted on a moving XYZ stage, allowing for alignment, as well as automatic

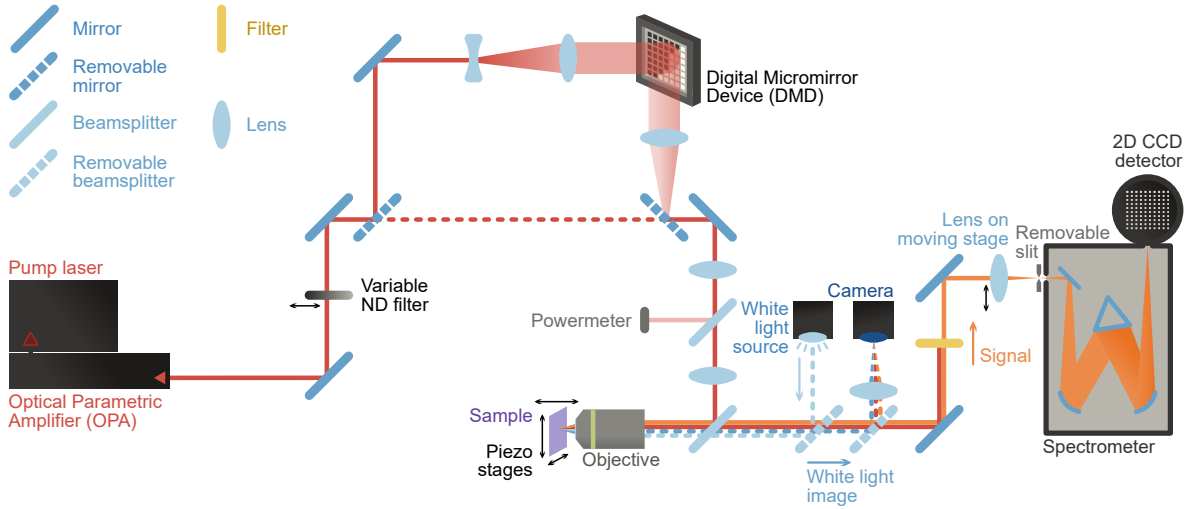


FIGURE 4.2: Diagram of the optical setup used for the lasing network investigation. The *Digital Micromirror Device (DMD)* can be inserted into the beam path using removable mirrors. The white light path is used for imaging and alignment and can be removed for measurement.

stabilisation for longer measurements. Removable mirrors allow us to insert a white light path, which shows the sample, pump, and signal on a camera, and allows for imaging and alignment.

For spectral measurements, a spectrometer is used to diffract the emitted light as before. Notably, a 2D *charge-coupled device (CCD)* camera is used instead of a line array, meaning that we are able to capture images that preserve one spatial axis (the vertical ‘y’ axis) that is not diffracted. I will refer to these as *hyperspectra*, to contrast with the 1D *spectra* seen before. A notable limitation of this approach is that our light sources (the networks) are extended in space, and so the diffracted axis (‘x’) will not in general correspond *purely* to wavelength – that is only true for point sources. For example, the emission from the left side of the network will be shifted slightly to the left of the camera’s x axis, and thus its ‘wavelength’ reading biased. For any measurements where knowing the exact emission wavelength is important (like Fig. 4.1 c) we thus introduce a removable slit in front of the spectrometer. The image from the sample plane is projected onto the narrow slit, and thus cut into a ‘point source’ along the x axis, negating any spatial smearing. The drawback of this is that we only see the signal from a narrow slice of the network defined by the slit, and so for most measurements where the precise wavelength is not critical we remove the slit and accept some spatial information being mixed into the wavelength axis.

Armed with a detailed understanding of the experimental conditions, let us now have a closer look at the spectral features of a lasing network. Fig. 4.3 a shows a full hyperspectrum from a bigger, 150 μm -diameter lasing network. This one is taken with the slit removed, and we can see

that the spectral landscape is exceedingly complex – many broad peaks coming from spatially extended regions blend together to create a non-uniform blur of features across the InP gain spectrum.

Thanks to the spatial axis of the camera, we can look at specific ‘cuts’ of the network in the y direction, shown in Fig. 4.3 **b**. We can see the individual lasing peaks more clearly here, but each row of pixels will still be a mix of the output from many different nodes of the network, and so the spectra are still mostly a Gaussian-shaped collection of many peaks. To go a layer deeper, we can insert the slit in front of the spectrometer to constrain the x axis to a specific vertical slice of the device. This way we constrain both spatial axes – x with the slit, and y with the chosen camera row. The lens that projects the image onto the slit is mounted on a moving stage (see Fig. 4.2), and so we can change which slice of the image goes into the spectrometer, and thus select specific points of the sample to measure spectra from.

Fig. 4.3 **c** shows the spectra from three selected nodes of the network (the inset highlights

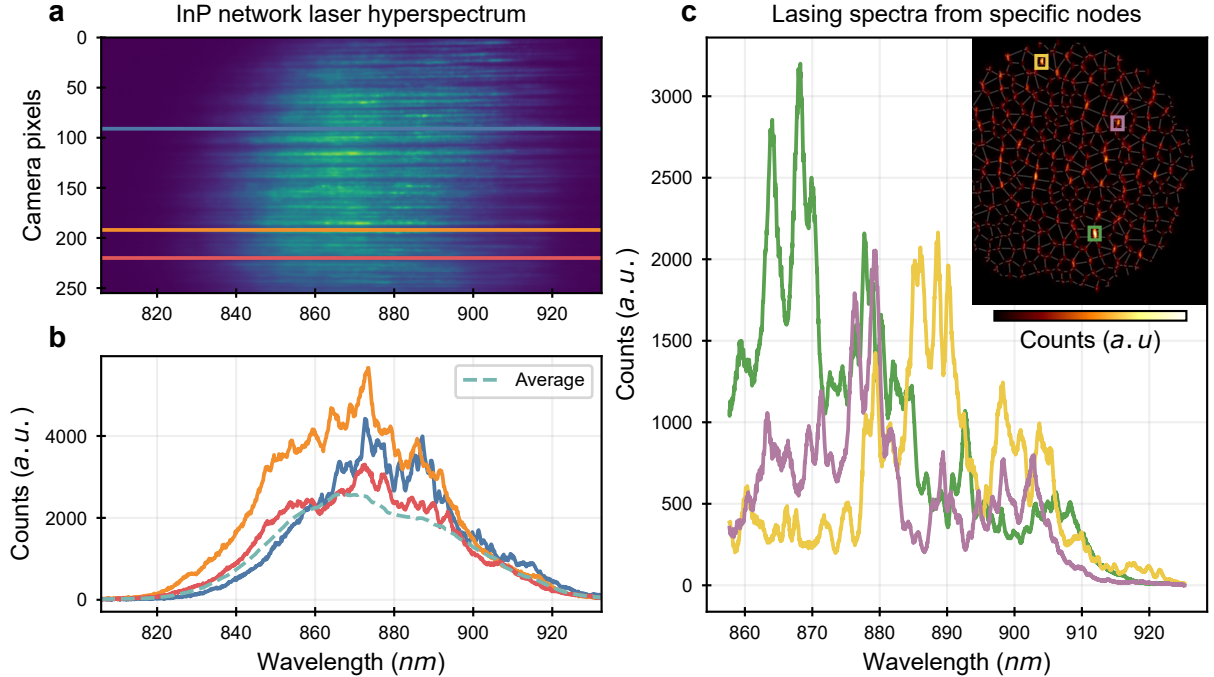


FIGURE 4.3: The spectra measured from the network lasers are complex and intermixed. **a** A hyperspectrum emission from an InP lasing network 150 μm in diameter, with 350 nm wide waveguides. The image is diffracted along the x axis. **b** Three slices through **a**, indicated with horizontal lines in the hyperspectrum. The peaks in the spectrum are wide, and mix together to create a very broad lasing feature. **c** We can focus on specific network nodes to see the spectra more clearly. Here, three nodes are chosen, and their spectra overlaid. The node locations are shown on a far-field image in the inset, obtained by summing over the spectral axis of the slit-scan data. The shape of the network is overlaid in grey. A forest of broad, overlapping modes is seen, with each node supporting many modes, and each node’s emission being significantly different from the others.

them on a far-field image). The individual peaks can be seen more clearly, showing that dozens of modes pass through each node, closely-spaced in wavelength. The modes are also spatially constrained (they do not span the entire network) and so the mode landscapes in different regions of the laser will be completely different.

This illustrates how complex and rich a system the random lasing network is. The network studied here has over 650 nodes, and each outputs a forest of modes at different intensities and wavelengths, with each mode spanning across many nodes and edges. The ability to pattern the pump beam allows us to influence these dynamics too, delivering more or less gain to different regions of the network, which can have drastic effects on the mode landscape within.^[65] The rest of this chapter explores these complex dynamics and the principles behind them.

4.2 The Physics of network lasers

4.2.1 Transport within the network and coupling between the waveguides

We understand from Chapter 3 that lasing requires amplification and confinement. The source of amplification in our semiconductor networks is much the same as in the microdisks explored in that chapter – the device itself is made out of amplifying medium, InP in this case. Again similarly to microdisks, the device itself is also the source of confinement – in this case through multiple scattering at the network nodes, which keeps the light somewhat trapped in the labyrinth of the network. This scattering is unfortunately imperfect, and thus also the main source of losses in the network lasers. In this section I will explore these losses and outline some ways in which they can be minimised.

The impact of node shape on inter-waveguide coupling can be seen qualitatively in Fig. 4.4. Light above the material bandgap is coupled into a waveguide using a fibre coupler (illustrated in Fig. 4.4 **a**) and delivered to a system of network waveguides without amplification or absorption. Fig. 4.4 **b** and **c** compare how the light is transported through the network for two kinds of junctions.

First, a network similar in shape to the old electrospun polymer platform is studied in Fig. 4.4 **b**. Here, all nodes have a degree of four (the node is connected to 4 other nodes) as the nodes are formed from two long waveguides overlapping at a random angle, imitating fibres deposited through electrospinning. The losses at each node are comparatively low, but that is because for each node there exists a preferred, low-loss path forward, and the scattering to the sides is reduced. The light thus traverses deep into the network, but mostly along a smaller

number of straight paths.

Fig. 4.4 c shows the new, Voronoi-based network topology. The degree of each node here is three, and there is not a preferred path forward – the light has to scatter sideways to proceed. This leads to a wide spread of paths the light can take, but also higher losses. Unamplified light travels only a couple of nodes away from the input before being completely lost. We can thus see that the topology we choose for the network can impact the amount of loss, but also the kind of preferred travel paths for the light within.

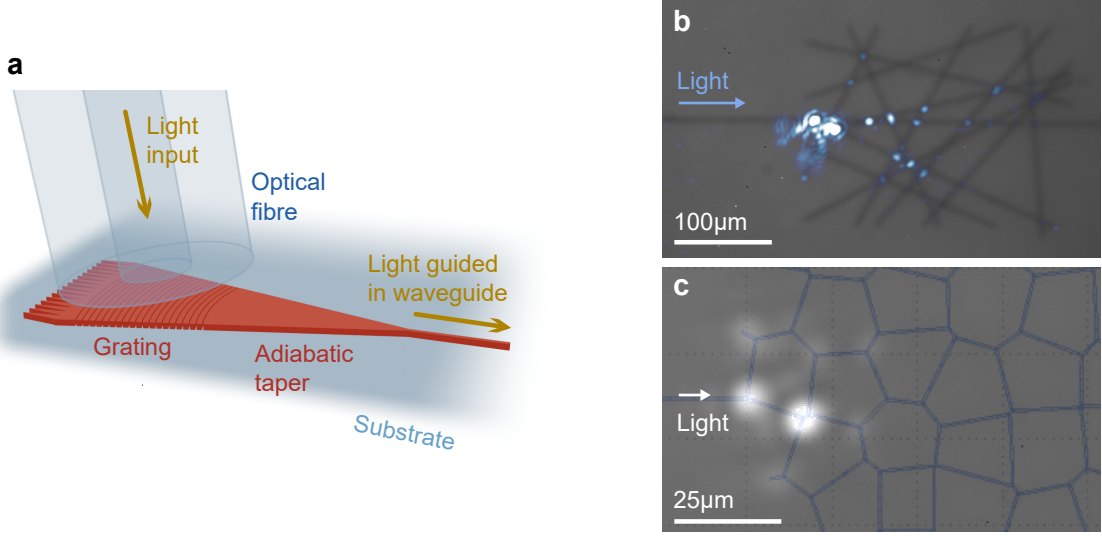


FIGURE 4.4: Studying light transport within the networks. **a** A schematic of the setup used to couple light into the waveguides for this test. Light is delivered through an optical fibre, coupled in from free space using a grating, and then delivered into a waveguide with an adiabatic taper.^[158] **b** Scattered light collected in the far field from a network design based on the electrospun platform. **c** Scattering from a network of the new, Voronoi-based design.

We can turn to simulations to understand the guiding and coupling of light within the network better. Initial FDTD simulations indicated that wide, lower height waveguides have lower scattering losses at nodes than narrow, tall waveguides,^[159] which worked well with the ~ 120 nm thick bonded InP wafers we had at the time. The rest of this study assumes a 125 nm layer of InP and varies the waveguide width.

Fig. 4.5 a shows a series of *Finite Difference Eigenmode* (FDE) solver^[160,161] simulations of the guided modes in waveguides of varying widths, sorted by their effective refractive index. The evolution of this effective index, as well as mode profiles for a more granular range of waveguide widths, can be seen in Appendix Fig. B.3. 350 nm waveguides primarily support one *Transverse Electric* (TE) mode with low losses (TE₀). As the waveguide width grows, the double-barrelled TE₁ mode becomes better-supported, swapping places with the dominant

Transverse Magnetic (TM) mode (TM0). While we were initially aiming for single-mode operation, which would better match the assumptions of our network mode simulation framework netSALT (discussed in Section 4.2.2), the substrate introduces anisotropy in the waveguide geometry and lifts mode degeneracy,^[162] and so the waveguides are multimode at all widths.

The mode shape becomes relevant once we look at the scattering patterns at the network nodes. We model a symmetric Y branch (degree 3 nodes with 120° angle between links) as depicted in Fig. 4.5 b and calculate the mode transmission using FDTD simulations (this simulation was performed by Dhruv Saxena). The transmitted mode power from one link into adjoining links at the node is split into all available guided modes (the ones simulated above). This can be seen in the field cross-section profiles in Fig. 4.5 b, where the field in adjoining links is not the same as that in the injecting link. The TE0 mode loses most of its field as it scatters, but the double-barrelled TE1 mode in the 450 nm waveguide intuitively ‘splits’ across the two branches, resulting in higher transmission. Ref. [159] contains a detailed study of transmission values and inter-mode coupling at a junction of this kind.

The practical implications of this width-dependent scattering can be seen in the thresholds of networks with different waveguide widths. Fig. 4.5 c compares the LL curves for two networks of the same topology but 350 nm or 450 nm wide waveguides. The scattering losses are lower for the

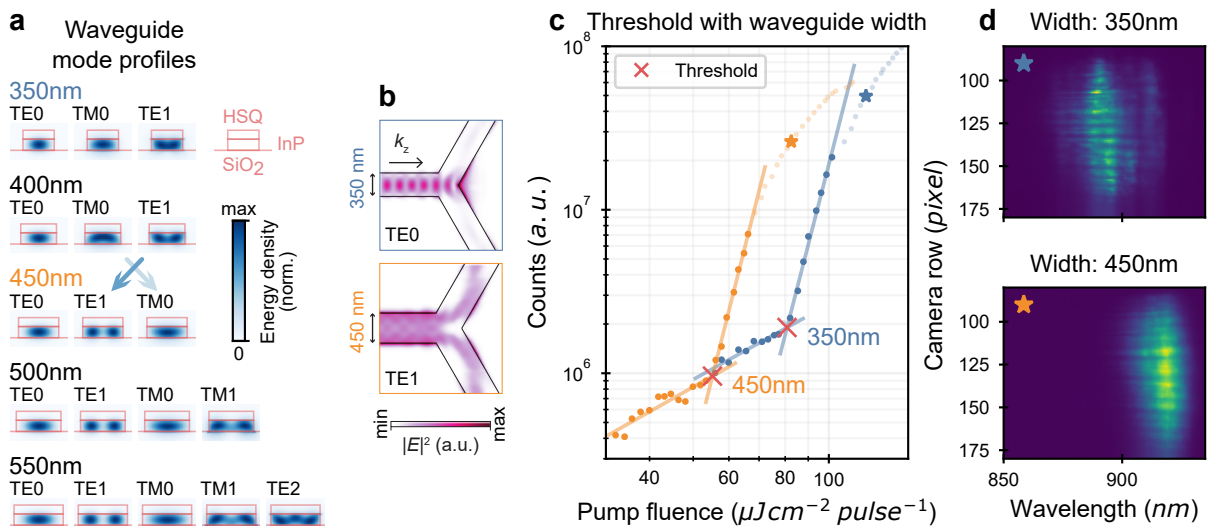


FIGURE 4.5: Waveguiding and waveguide coupling within the network. **a** Supported mode profiles inside of waveguides with different widths, sorted left to right by their effective refractive index, simulated with a FDE solver for light at 880 nm. **b** Behaviour of light at a waveguide junction simulated with a FDTD solver.^[124] **c** The lower losses at waveguide junctions result in a lower lasing threshold for the entire device, illustrated here by LL curves for two networks with the same topology, but different waveguide widths. **d** The hyperspectra for the two waveguide widths at 150% the threshold pump fluence are qualitatively different.

450 nm variant, which results in a significantly lower threshold pump fluence. This experimental relationship holds across network topologies.^[159]

Fig. 4.5 d compares the hyperspectral output from the networks at the two studied waveguide widths. The spectrum from the wider waveguides is notably narrower, and this is not entirely due to the increased blueshift at higher powers, as the relative narrowing persists when the pump fluence is comparable between the networks. The exact mechanism of this narrowing is not currently understood – the multi-modal nature of the waveguides could be changing the way the modes are established through multiple scattering, it could be a thermal effect due to the larger gain volume of the wider waveguides, or perhaps better node coupling allows for higher Q factor modes, which dominate the spectrum.

An additional source of losses that I have not considered here is the imperfect confinement of light within a single waveguide. Some losses will indeed arise as light is guided along, and the confinement factor as a function of width is simulated in Ref. [159] for our system. These losses are fairly constant though, and significantly smaller than those at the nodes. I have experimentally tested the guiding of light through etched waveguides around a millimetre in length, and no significant inefficiency was apparent even at that scale. Nevertheless, further optimisation of waveguide sidewall quality is a potential direction of study, especially in terms of reducing device degradation under prolonged illumination.

For the rest of this thesis most of the measurements shown will be from network lasers that use waveguides 350 nm wide. These were chosen in the end because they seemed to perform well for our application, had a very broad and rich spectrum, and the dominant TE₀ mode matched better with the assumptions made when simulating network modes. Despite the high losses at the nodes, the networks still support modes that are localised and coherent over long distances, and I will explore that more in the upcoming sections. The node shape and waveguide attributes remain very interesting parameters that could be tuned to obtain devices with lower thresholds, and thus reduced power requirements and less thermal degradation.

4.2.2 Lasing modes within a random laser

We have established the large number of lasing modes that a network laser can support (cf. Fig. 4.3). Let us now explore some tools to understand the nature of these modes better, first through theory, then through experiment.

NetSALT^[63,65,163] is a simulation suite which we can use to analyse the lasing network. First, the network is treated as a graph, and wave equations are solved along the graph's

edges to establish the light modes that the structure supports passively (see the Supplementary Information of Ref. [63] and Ref. [65] for a detailed treatment). Then, the evolution of the modes in the complex $k = k_r + ik_i$ (wavenumber) plane is computer and tracked under the *steady-state ab initio* (SALT) approximation^[66,164] as the pump power is increased (see the Supplementary Information of Ref. [65] for details). The passive modes start with negative k_i values (indicating the mode is lossy), but at higher pump powers some of them will reach positive k_i , indicating lasing.

Fig. 4.6 shows electric field intensities for three network modes simulated with netSALT by T.V. Raziman. The modes are localised to different areas of the network, with each predicted to span many links. The full simulation predicts 239 passive modes, with at least 172 lasing, and the degree of localisation varies between them. The first important takeaway is that this system will lase as a complex whole, rather than a set of independently lasing Fabry–Pérot cavities. The modes are also predicted to overlap spatially, which would not matter in a linear system, where light waves do not interact with each other, but matters a lot in a non-linear system, where lasing modes can deplete the available gain. This will be discussed in detail in Section 4.2.3.

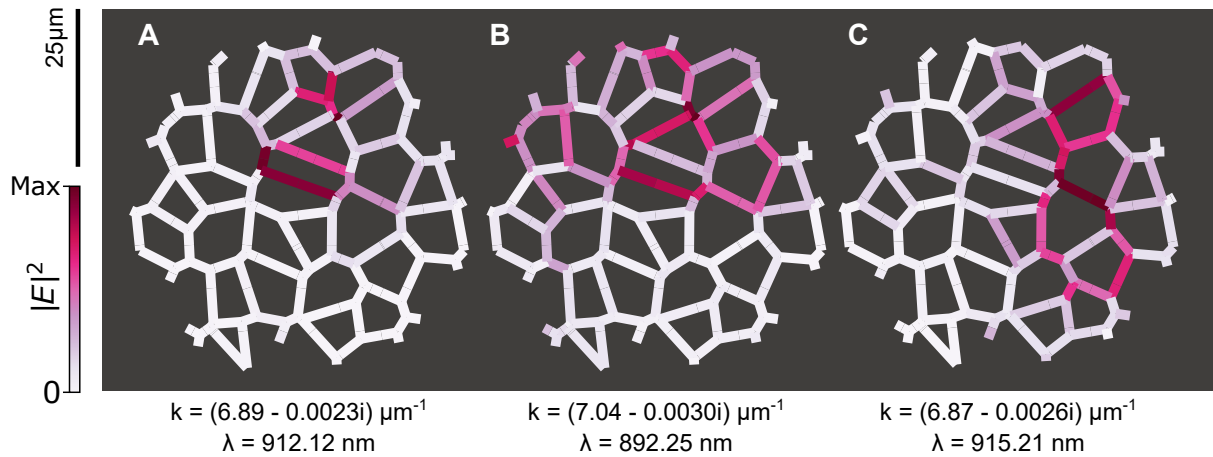


FIGURE 4.6: A selection of 3 passive modes for a 50 μm lasing network, simulated using the netSALT software. The complex wavenumber is given, indicating the mode wavelength and passive losses.

While netSALT predicts many lasing modes that span large areas of the network, it does not currently take into account light losses at the nodes, which we have seen play a large role in Section 4.2.1. Supporting these large modes requires long-range coherence for the phase matching conditions to be met, and we have seen that unamplified light is lost after only a few waveguide junctions. The amplification from InP should enable coherent light to travel much further, but we would like to experimentally see these large modes to prove that this is indeed the case.

For this test we turn to the slit-scan technique used to obtain emission from specific nodes

back in Fig. 4.3. By shifting a lens immediately in front of the spectrometer slit, I could acquire a spectrally-resolved series of images corresponding to each slice of the network. From these, I have extracted a spectrum for each of the network nodes (like the three nodes shown in Fig. 4.3 c) and identified the lasing peaks within it. I then went through all pairs of spectra for nodes joined by a waveguide and matched up the peaks that I believed corresponded to the same lasing mode. Going edge by edge and matching the corresponding peaks, one can build up a sub-graph of all peaks belonging to that mode. This assumes that the nodes are contiguous – if the network emits light at a similar wavelength at two locations in the network, but nowhere in between, I assume these to be separate modes, as there can be no phase matching if light at that wavelength does not travel between the regions. Each of the resulting sub-graphs within the lasing network graph thus corresponds to a contiguous region of nodes that emit light at a similar wavelength. A selection of these is shown in Fig. 4.7.

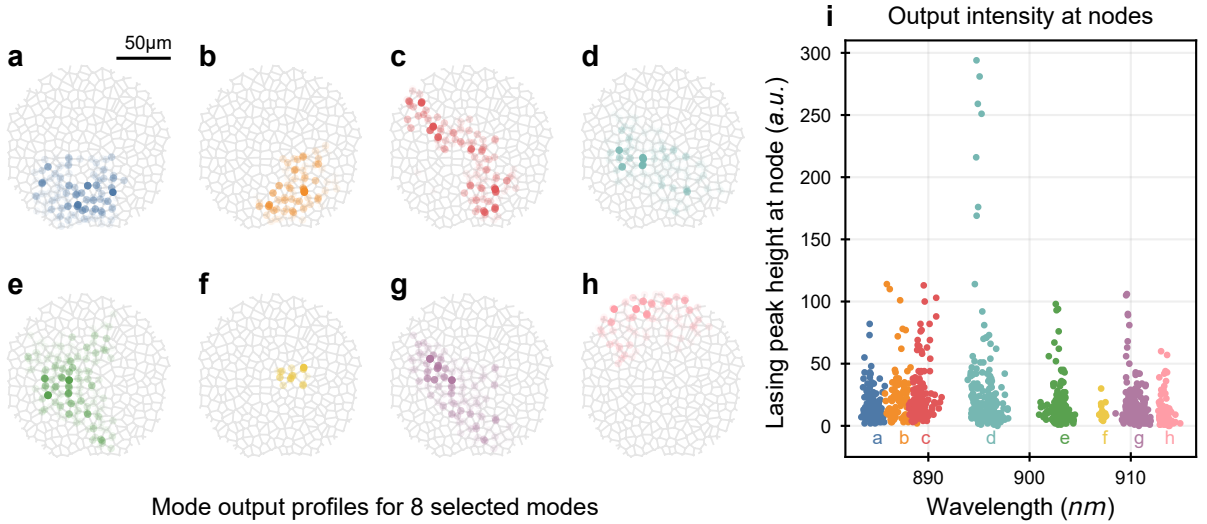


FIGURE 4.7: Selected mode profiles measured from a 150 μm lasing network. **a-h** Nodes where signal from the chosen mode is seen, with circle opacity corresponding to the measured intensity of the mode at the node. **i** Intensities of the nodes highlighted on the left plotted in terms of wavelength, showing the λ separation between overlapping modes (like *c* and *g*).

There are some caveats to note here: the peaks are generally difficult to analyse and pair up, as they are broad, often close together, and varying blueshift plus experimental uncertainty can make it difficult to discern which mode a particular peak belongs to. The investigation yielded over 200 potential ‘modes’, but care needs to be taken when analysing them, as some of these could be multiple modes joined together, and some could be one mode accidentally split apart. This data also only shows where the mode outcouples, and not necessarily its phase-matched path – the outcoupling is likely to come from a larger area as the mode randomly scatters into

neighbouring waveguides.

We can still draw some interesting conclusions from the data. First of all, it is clear that despite the high scattering losses and short unamplified transport, long-distance coupling prevails when light is amplified, with modes spanning large regions of the 150 μm network. The degree of localisation varies, with modes like **f** spanning only a couple of links, and modes like **c**, **e**, and **g** going nearly edge to edge. Different modes are localised throughout the network, including at the edges like **h**. The losses from the waveguides ending there may be higher than in the middle (where **d**, the strongest mode, lives), but modes are still supported. Finally, multiple modes at different wavelengths can clearly co-exist in the same region, sometimes with a little overlap (like **a** and **b**), and sometimes with a lot of overlap (like **c** and **g**).

For further detail on mode localisation, Ref. [159] contains an independent experimental treatment that also verifies that the waveguides are lasing as a system rather than a collection of small cavities.

4.2.3 Lasing mode competition

When multiple modes exist in the same area, they will all try to antagonistically use up the available gain for their own stimulated emission.^[164,165] We refer to this as *mode competition*. This can result in highly non-linear behaviour, as stronger modes can suppress the weaker ones, and the weak modes dying frees up more gain that can further enhance lasing from others.

An experimental illustration of the complex dynamics that arise is shown in Fig. 4.8. As the pump power delivered to the laser grows, additional modes can reach threshold, upsetting the existing mode balance. As a result, we see specific spectral regions of the hyperspectrum plateau or even decrease in intensity, sometimes multiple times at varying rates. This is in stark contrast to the usual linear relationship of lasing output with the input power. Each mode will have a slightly different non-linear behaviour, coupled in a non-trivial way to the modes that it shares physical space and gain with.

The netSALT model introduced in Section 4.2.2 includes the competition between spatially-overlapping lasing modes and so we can use it to probe this dynamic further. Specifically, we will see how delivering more or less gain to specific regions of the network can enhance or suppress the modes in that region, and the implications this has for the rest of the network.

Fig. 4.9 **a** shows three simulated spectra of a lasing network (simulations performed by T.V. Raziman). The spectra correspond to three illumination patterns: uniform illumination (‘full’), horizontal edge (‘right’) and vertical edge (‘top’) (shown in Fig. 4.9 **b**). We will focus on three

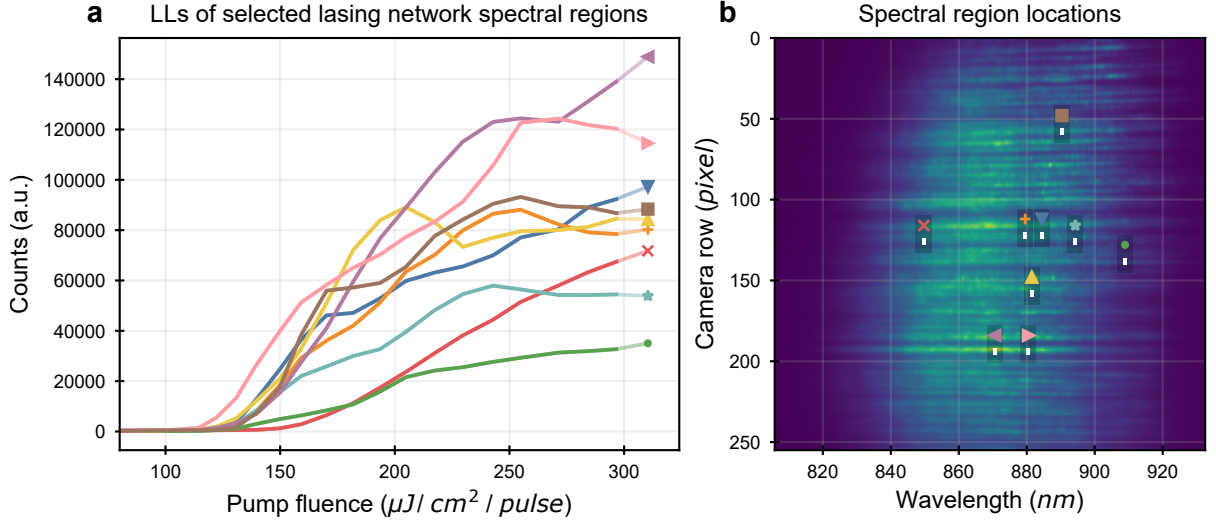


FIGURE 4.8: Modes in a network laser behave non-linearly. **a** Intensity changes for selected hyperspectral regions as the pump fluence increases. The region positions along the spectral axis were compensated for blueshift at each power. The intensities behave non-linearly, exhibiting different thresholds, and different evolutions beyond the threshold. **b** A map of the chosen hyperspectral regions at the highest pump fluence.

specific modes, labelled A, B, and C, identified in figure 4.9 a by dashed vertical lines. The profiles of these modes were previously seen in Fig. 4.6, and are reproduced at the bottom of Fig. 4.9 c for reference.

Fig. 4.9 c shows the lasing intensity of the three modes under ‘full’, ‘right’ and ‘top’ illumination patterns. Under full illumination, mode B shows the highest intensity, while mode C does not lase at all, as it falls below the lasing threshold. Mode B is mostly localised in the top area of the network though, so under ‘right’ illumination it loses a lot of its gain and stops lasing. Mode C, which overlaps B spatially, can use the gain freed up by it and go above threshold for this illumination. For a ‘top’ edge illumination mode C once again disappears, with the competition between A and B playing out differently than under ‘full’ illumination – mode A is now much stronger.

Notably, some of the modes lase more strongly under partial illumination than under full illumination. This is unusual, and a sign that the mode competition influences the behaviour of the network strongly – we put less energy in the system overall, but get *more* output from some parts of it. This strong preference of certain modes for specific projection patterns will become relevant in Chapter 5, where we use it for feature detection in a machine learning scheme.

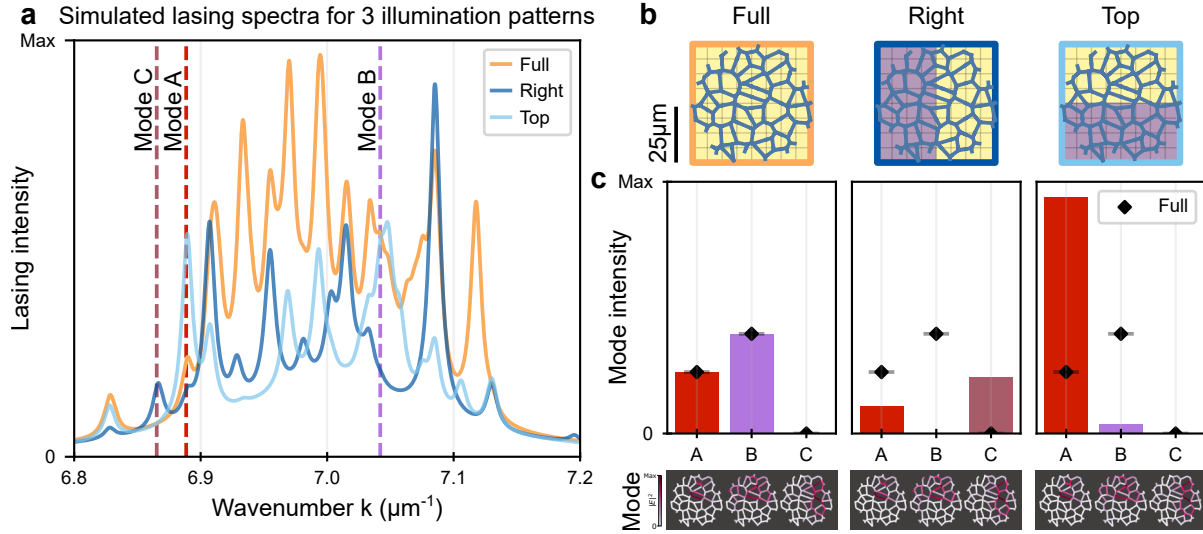


FIGURE 4.9: The principle of mode competition. **a** Lasing spectra of a 50 μm diameter network under three illumination patterns, simulated with netSALT. The wavelengths of three modes (A, B, and C) are highlighted with dashed lines. **b** The three illumination patterns overlaid on the shape of the network. **c** The intensity of the modes ABC under the three illumination patterns. Mode shapes from Fig. 4.6 are reproduced below for reference. The illumination shape changes the intensity of the modes, with some decreasing intuitively with respect to their full illumination intensity (highlighted with a black dot), but some increasing under partial illumination.

4.2.4 Long-distance coupling and disturbing it

We have seen that changing the illumination pattern in one network region can influence the modes in a different region through mode competition, and that there is long-range coherence in the network as the modes are spatially extended. But how far does this coupling extend, and can we understand it better by disturbing it?

Fig. 4.10 shows an experimental study where the top half of the network is constantly illuminated, but the illumination of the bottom half varies from fully illuminated to illumination under the lasing threshold. The spectrum of the top half of the network is shown in the top row of Fig. 4.10 b, and we can see that it changes as the illumination of the bottom half goes from ‘on’ to ‘off’, with some parts of the spectrum being enhanced, and some suppressed. We can also tune the illumination of a separator region between the top and bottom halves of the network, and observe how the magnitude of the coupling (bottom row of Fig. 4.10 b) increases as we give that region more gain.

This shows that there is long-range coupling in the network, with the bottom half influencing the top half, but also that we can shape and suppress that coupling in part *thanks* to the high losses at the network nodes and in the unpumped material. The high losses make patterned illumination impact the viability of specific lasing modes significantly, and this impact is then

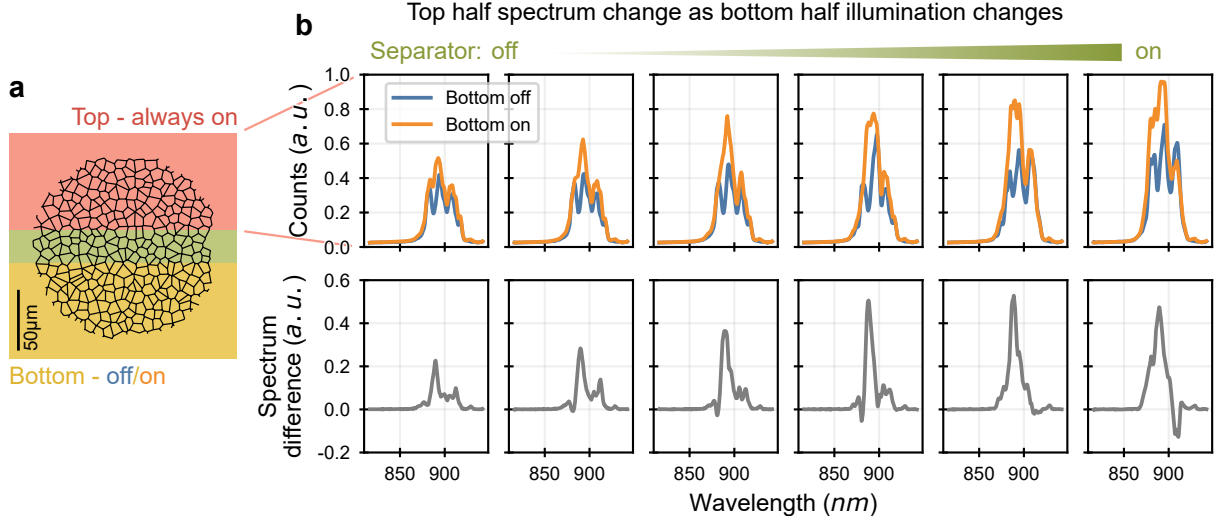


FIGURE 4.10: Losses disturbing long-range coupling in a $150\ \mu\text{m}$ diameter network. **a** Illustration of the illumination layout for this experiment. The top section of the network is fully illuminated, the bottom half is switched from full illumination to below threshold (‘off’). The middle section is smoothly tuned from ‘off’ to fully illuminated. **b** A spectrum is collected from the top half of the network only. The emission from the top half changes as the illumination of the bottom of the network switches from ‘off’ to ‘on’ (top row). The magnitude of this change (bottom row) increases as the separator section brightness raises, showing improved coupling.

further amplified by mode competition interactions. Depending on the desired application, one may thus want to tune the network topology to magnify losses, increasing the impact of the pump’s shape on device dynamics, or decrease the losses to promote long-distance cross-talk.

4.3 Application considerations

4.3.1 Sensitivity and resilience

The spectrum of a network laser is extremely sensitive to the illumination pattern thanks to the long-distance mode competition discussed in this chapter, as well as effects like mode reshaping.^[65] This kind of sensitivity can be used to create a tunable multi-peak laser,^[65] sensitive detectors, signal processing, or for machine learning applications (as discussed in Chapter 5). Fig. 4.11 **a** shows an example of this sensitivity for a more complex image. Pumping the network with a spatially slightly different pump beam can be seen to highly influence the emission spectrum.

For a real application, the output of the laser has to be not only sensitive, but also stable over time. The InP network lasers show a high degree of stability towards degradation, explored in Fig. 4.11 **b-d**. We studied the lasing stability of a network with $150\ \mu\text{m}$ diameter, $7\ \mu\text{m}$ average

link length, and 550 nm waveguide width under prolonged pumping. The network was pumped continuously with a 633 nm, 225 fs-pulsed laser at 100 kHz for 1.5 hours, and spectra were acquired every 5 seconds. Total emission intensity of these spectra decreased only slightly ($\sim 20\%$) over 4.5×10^8 pump pulses (Fig. 4.11 b). The lasing threshold of the device only increased by 3% after performing this intense pumping regime, indicating little permanent damage (Fig. 4.11 c), and the shape of the spectrum remains similar (Fig. 4.11 d).

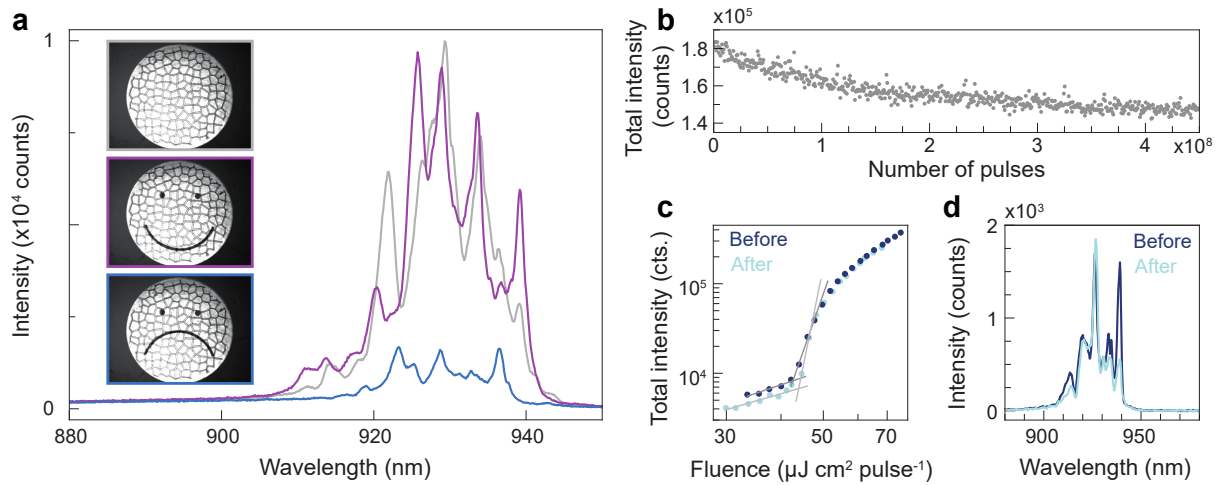


FIGURE 4.11: Spectral sensitivity and emission stability of InP network lasers. **a** Pumping a 150 μm diameter, 550 nm waveguide width network with different spatial patterns: full circle (grey), happy face (purple), and sad face (blue) strongly affects the emission spectra (taken with slit inserted). **b** The emission intensity of the network over many excitation pulses. The total emission intensity decreases by 20% over 4.5×10^8 excitation pulses. The device was pumped continuously with a 100 kHz laser and a spectrum was acquired every 5 s. **c** LL curves and **d** spectra of the network before and after the stability test in **b**. The threshold and spectrum remain comparable after the prolonged illumination.

Our InP network laser is thus much more stable than polymer-based devices, and a fitting platform for the machine learning experiments coming up in Chapter 5, which require the system to be illuminated for many hours or days at a time.

4.3.2 Network design and topology

One of the core benefits of this new DWB network laser platform is that the networks are freely designable. Controlling the shape of the network could let us tune the spatial overlap and mode extent and thus the degree of coupling and non-linearity, reduce the losses in the system, and chose exactly how we break the symmetry of the cavity.

The Voronoi networks used throughout this thesis were chosen as a reasonable starting point. They start off as a regular (square or hexagonal) grid of seed points, which are then displaced

randomly. This avoids clustering of points (inherent to completely random point allocation), resulting in semi-uniform waveguide lengths. A Voronoi diagram is then generated for the points – this is a space partition, where each cell consists of the areas of a plane which have the same nearest random seed point. We then crop the Voronoi diagram to the desired network diameter, and the edges between the diagram’s cells become our waveguides.

This approach has several advantages. Each node consists of exactly three waveguides joining, the angles between them tend to not be extreme, and the degree of disorder and waveguide lengths can be smoothly controlled by tuning the original point distribution. Changing the random generator’s seed parameter also allows us to create sets of networks with similar overall properties but different waveguide arrangements. The main disadvantage of this design are the high losses at waveguide junctions discussed in Section 4.2.1.

Other network designs are a very interesting area of potential exploration. We have generated networks with varying degrees of disorder to study the impact of complexity on the lasing, networks with a fixed node shape but varying edge lengths to minimise losses, networks that match the polymer-based platform shape better, and variants that explore the boundary conditions at the network’s edges. Fig. 4.12 shows that networks fabricated based on various designs do lase reliably, proving the high flexibility of this platform. There is thus much left to explore in terms of geometries that may work better for specific tasks and applications.

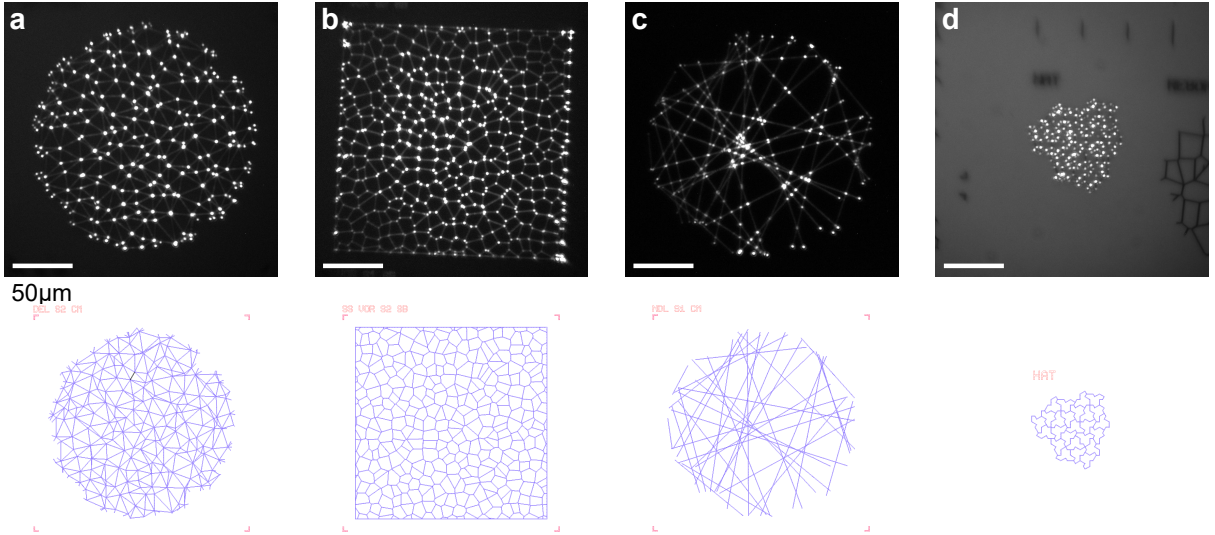
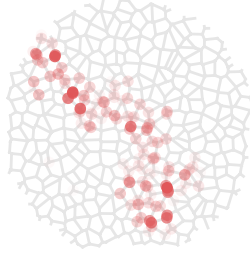


FIGURE 4.12: Far-field images of the lasing emission from four examples of network topologies, demonstrating that working network lasers can have a large variety of shapes. The bottom row shows the computer-aided design for the devices above. Emission is primarily from the network nodes, which is characteristic of lasing.^[63] **a** Delaunay triangulation of semi-regular points. **b** Voronoi graph with a square border. **c** Randomly oriented lines designed to imitate the electrospun polymer-based networks.^[63] **d** A design based on the ‘hat’ aperiodic monotile.^[166]

Conclusions

This chapter shows network lasers to be a complex, highly non-linear system. We demonstrate that there is a high number of spatially-extended, overlapping modes that they can support, and that these modes interact with each other, and respond to patterned illumination in highly non-linear ways. We show that losses and mode competition play a role in shaping the behaviour of the system. Finally, we demonstrate that the networks are resilient to long-term illumination, and can lase while having wildly varying topologies. Together, these features make them a promising platform for neuromorphic computing, and I shall explore this application of the network's physics in the next chapter.



Chapter 5

Neuromorphic computing with network lasers

For the work presented in this chapter I have contributed most measurements, development of the experimental edge detection concept, and parts of the data analysis with assistance from J. C. Gartside. Most figures in this section, as well as parts of the relevant discussion, are adapted from the following publication, plotting for which has been done by me, A. Fischer, and W. K. Ng:

J. Dranczewski,* W. K. Ng,* A. Fischer,* T. V. Raziman, D. Saxena, T. Farchy, K. Stenning, J. Peters, H. Schmid, W. R. Branford, M. Barahona, K. Moselund, R. Sapienza, J. C. Gartside, “[Retinomorphic Machine Vision in a Network Laser](#),” *arXiv*, Aug. 2024. [*manuscript in review*]

*Authors contributed equally.

Abstract

Semiconductor network lasers are compact, highly non-linear systems. We thus believe that they can provide potential solutions to the problems of neuromorphic photonics, and in this chapter I will explore their application to a number of computational tasks. First, feature detection is demonstrated, a task closely matching the mode competition dynamics within the laser. We detect black and white image edges by scanning a small window of it across the network and observing mode ‘activation’, where specific spectral regions lase more strongly than under full illumination. A feature detection dataset also allows us to experimentally demonstrate suppressive mode action, and thus negative weights. Next, we use the extreme learning machine architecture to perform image classification, including MNIST (digits), Fashion MNIST (clothing), and BreaKHis 400X (breast cancer histopathology). By combining feature detection and image classification into a two layer, [CNN](#)-like scheme, we demonstrate further improved performance,

especially when training data is scarce, and compare our results to purely digital benchmarks. Finally, we explore a number of potential explanations for our strong neuromorphic performance.

5.1 Feature detection

Given the strong response of the networks to spatial patterning of the pump beam, the first application I will explore is feature detection, where we project a simple image onto the network, and are interested in whether that image is a ‘feature’ – for example a left or right edge of a bright object. As introduced in Section 1.1, *Convolutional Neural Networks* (CNNs) are a powerful neural network architecture for image processing which start with a number of feature detection layers like this, using the process of convolution illustrated in Fig. 1.2 c.

The computational intensity of convolution makes it a strong target for physical hardware implementations. Recent examples of convolutional hardware processing include memristor arrays,^[33] photonic integrated circuits,^[44,167] spiking VCSEL lasers,^[30] and meta-surfaces.^[41] These schemes largely adopt the structure of software CNNs, implementing physical equivalents of the linear matrix multiplications.

5.1.1 Lasing mode competition for convolution-like feature detection

We believe that the mode competition in our system maps to the challenge of feature detection in an interesting way. In Section 4.2.3 we have seen that projecting specific patterns onto a network laser can shape the mode competition such that modes that were previously weak can now lase strongly. I will refer to a mode lasing more strongly than under full network illumination as mode ‘activation’. Another distinction from the theoretical study is that we cannot directly inspect the mode strengths within the laser, and so when referring to ‘modes’ in the experimental context I mean regions of the hyperspectra produced by our spectrometer. These regions may have multiple lasing modes mixed within them from different spatial regions along the x axis.

Fig. 5.1 shows how we can use this mode activation to experimentally detect features in an image. Our example image, a photo of a cameraman, is shown in Fig. 5.1 a. We split it into 4x4 pixel windows (example shown in inset), similar to how a raster-scan in convolution sequentially looks at small parts of the image (cf. Fig. 1.2 c). Three example image windows are displayed in Fig. 5.1 b, shown with their respective network lasing spectra and mode intensities for six modes. The lasing intensity changes for different illumination patterns – the overlap between the input image shape and spatial mode profiles changes for different patterns, and the modes non-linearly

compete with each other for the patterned gain. We observe that specific modes lase strongly when a given image feature (e.g. a left or a right edge) is present, providing feature detection via the lasing amplitude. Fig. 5.1 c illustrates this by showing the intensity of a selected mode when image windows from different places in the cameraman picture are illuminated on the laser. In fully dark areas the mode will produce no output (blue in the figure), and in fully white areas the mode will be as bright as under full network illumination. We consider a feature ‘detected’ when a mode ‘activates’ as defined above (highlighted red in the figure) – its intensity is higher for a partial illumination than it would be under full illumination. By plotting a spatial intensity map like Fig. 5.1 c for all the modes in our hyperspectra, I was able to select a number of ‘feature detecting’ modes that exhibit this activation behaviour. Figure Fig. 5.1 d shows activation maps for six of the chosen modes, with different colours corresponding to the six modes indicated in the hyperspectra in Fig. 5.1 b.

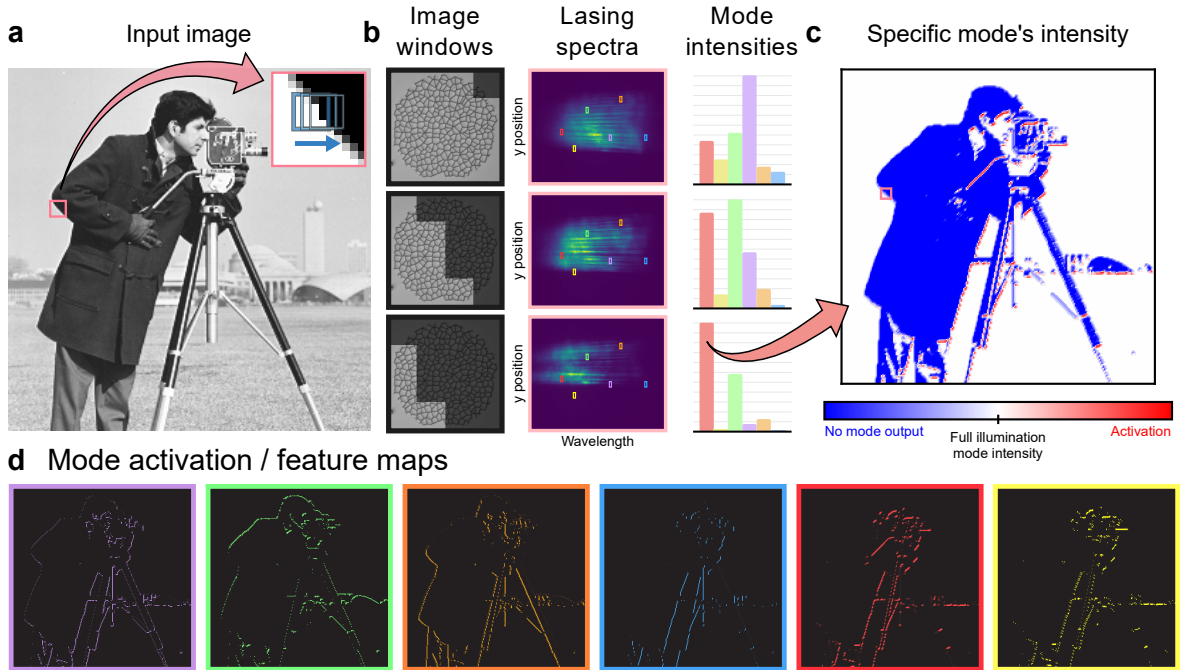


FIGURE 5.1: Convolutional processing through a complex network laser. **a** The input image (Ref. [168, 169], CC BY-NC 4.0) is divided into small windows, each containing an $N \times N$ pixel slice of the image (blue boxes in inset, here 4×4 pixels). **b** The image windows from **a** are binarised (see Appendix Fig. B.4), and illuminated on the InP network. The corresponding hyperspectral output is shown on the right, with mode intensities displayed as a bar chart for six feature-detecting lasing modes. Mode amplitudes vary in response to the presence of different image features. **c** We can plot how a selected mode’s intensity varies for image windows taken from different locations of the original image. Some image windows cause the mode to ‘activate’, with its output stronger under partial illumination than under full illumination (highlighted here in red). **d** Plotting only the mode activation from **c** gives us feature maps, indicating the regions of the image that the mode responds to strongly. Six maps are shown here, for the modes indicated on the hyperspectra in **b**.

A broad range of features is detected, with separate outputs for each feature provided by the different spatial and spectral locations of the modes. The detection is done in parallel – a single image window scan can produce multiple feature maps from different modes. This is in contrast to CNNs and their physical implementations, which usually have to detect multiple features through separate raster scans, and sometimes different hardware devices specific to a given kernel.

Because the process is based on mode competition *and* lasing, it is also inherently non-linear, with the modes corresponding to different features coupled with each other in complex ways. This is quite unlike the independent, linear kernels of usual CNNs,^[20] but is quite promising for machine learning applications, where non-linear activation plays a crucial role.^[18]

Fig. 5.2 explores this process in more detail, with Fig. 5.2 **b** showing how a hyperspectrum from an example 17x17 partial patterned illumination (image inset) differs from that of a fully illuminated network (Fig. 5.2 **a**). Most regions of the hyperspectrum are darker, but some are ‘activated’ – brighter than under full illumination, indicated in red.

Four modes are selected from the hyperspectra, with their intensities as the cameraman image is scanned across the network plotted in Fig. 5.2 **c**. The modes highlight certain features – bright-white edges can be seen in different locations depending on the selected mode. The reference full illumination region is indicated with a star. We can see that the mode output is brighter for the activated edges than in the fully illuminated (white) regions.

Fig. 5.2 **d** shows a combined activation map for the four modes selected here. Thanks to the large size of the image windows we are working with here, an edge reaches different parts of the network at multiple locations as it is swept across it. In this example, we see that the orange and blue modes highlight similar edges but with some offset. This indicates that the detecting modes are localised in different regions of the network, and activate at different times as the image is swept across. We believe that it is thanks to this localisation that there is a strong coupling between the structure of the input image and the hyperspectral output it produces.

The 17x17 image window used for the above study also illustrates the flexibility of our setup. Hardware implementations of convolution are often fixed to process image windows and kernels of a certain size, matching the number of interferometers in the system for example.^[44] We can instead process image windows of near-arbitrary size, limited only by the resolution of our DMD and the ‘resolution’ of the lasing network (the number of distinct edges and their distribution).

Fig. 5.3 showcases our capabilities across raster window sizes, illustrating the range of high-to-low spatial frequency features detected as kernel size is increased. Using multiple kernel

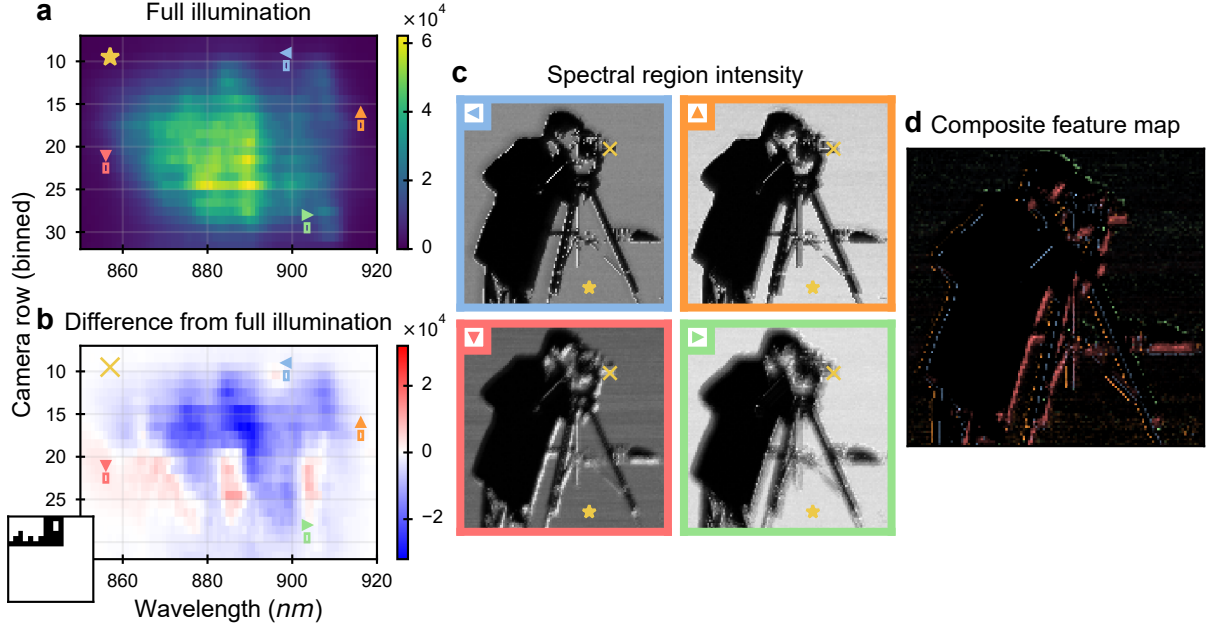


FIGURE 5.2: Exploring feature detection in detail. **a** Hyperspectrum of a $150\ \mu\text{m}$ diameter network under full illumination. This spectrum will be the reference point we compare the other hyperspectra to. Its location in the raster scan is indicated with a star in **c**. **b** The difference between an example partial illumination spectrum and the full illumination hyperspectrum (from **a**). This spectrum's location in the raster scan is indicated with a cross in **c**. Projected pattern in inset. **c** Mode intensity maps for the spectral regions indicated with rectangles in **b** when a 17×17 kernel is scanned across the black and white version of the ‘cameraman’ test image seen in Fig. 5.1 *a*. **d** Combined activation maps for the four modes from **c**.

sizes in CNNs has lead to state-of-the-art models^[170,171] but is computationally expensive.^[172] In our scheme, there is no additional computational cost to using larger raster windows, and the freedom to choose their size could enable interesting multi-kernel-size architectures. Large kernels have also been explored as a path to high performance in efficient networks,^[173] and our physical system is uniquely positioned to process these.

Fig. 5.3 also shows the results of our experimental scheme for colour image processing. Each of the *red*, *green*, and *blue* (RGB) channels of the image is illuminated separately, and several modes are selected per channel to create the composite maps shown in the figure. Each channel can be seen picking up distinct features, like the red nose or the blue cheeks of the baboon, which would be missed when considering only the intensity channel of an image. The separate illuminations are necessary due to the black-and-white nature of the DMD, and the network not being significantly sensitive to the wavelength of the pump light.

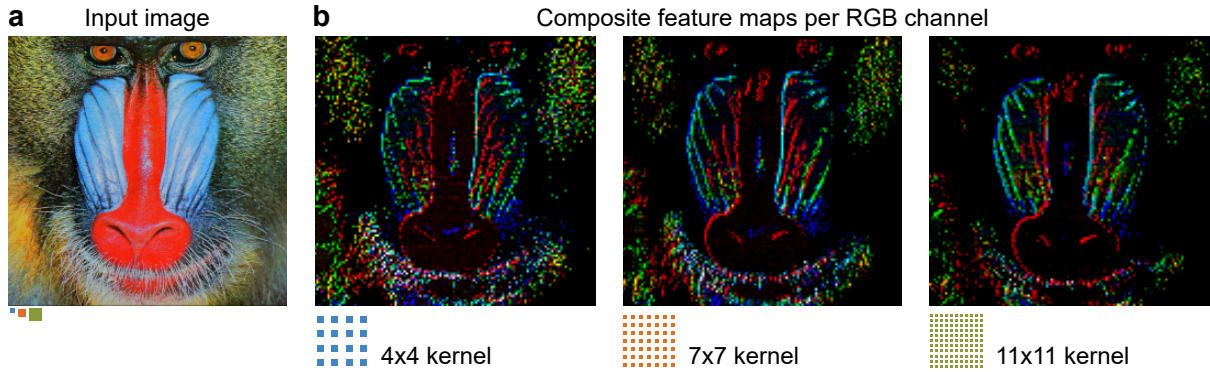


FIGURE 5.3: RGB edge detection capabilities across raster window dimensions. **a** The ‘baboon’ test image used for this test (public domain). Small squares below the image represent the three sizes of the raster window tested here. **b** Composite feature maps of the ‘baboon’ test image obtained with 4x4, 7x7, and 11x11 raster window sizes, combining multiple modes per RGB channel.

5.1.2 The 4x4 image library

While our feature extraction scheme is parallelised in the sense that multiple features can be extracted at the same time, it still requires sequentially acquiring all of the raster scan windows. Given current experimental constraints, we can only do this at a rate of around 10 Hz to 100 Hz. A scan like these shown in Fig. 5.3 for a 256x246 image consists of 14 514 exposures (per colour channel), taking a non-insignificant amount of time. Avenues for speeding up the measurement further are discussed in Section 6.2.1, but for the purposes of our current research we wanted to have a way to explore the feature detection mechanism of our network without needing to repeat lengthy measurements for each image.

If we constrain ourselves to a 4x4 raster window and binary images (matching the binary nature of the DMD), I have realised that there are only 65 536 possible images we would have to illuminate. I could thus create a library of hyperspectra matching these images, and refer to that to reconstruct experimental feature maps for arbitrary binary images digitally. Fig. 5.4 **a** shows the mathematical principle of constructing all possible 4x4 black and white images – a binary number is incremented one by one, and folded to a 4x4 grid to form the image. This gives each 4x4 image a unique identification number too, making the look up of specific images in the measurement data simple.

To select interesting modes out of this data, I have run the library look up on a test image of the letters ‘CL’ (for ‘Complex Lasing’), shown in Fig. 5.4 **d**. The image was split into raster windows like in the usual process, but instead of illuminating these on the laser, a hyperspectrum from the library was assigned to each raster window. This produced an activation map for each mode in the measured hyperspectra, and I could browse through these to select the ones that

picked out relevant features (six examples shown in Fig. 5.4 d). Fig. 5.4 b shows the locations of the 9 selected modes on the full illumination hyperspectrum taken from the library. This process is also how the edge maps in Fig. 5.1 were generated.

Having a hyperspectrum for all possible 4x4 images also allows us to run statistical analysis on the laser’s response to try to understand it better. To compare our physically-detected feature maps against linear matrix kernel filters as used in software CNNs, I have extracted effective kernel filters for our modes via least-squares fitting between the library’s input images and the measured mode intensity. These are shown in Fig. 5.4 c. The fits tell us which pixels enhance and which suppress a specific mode – it is notable that turning some of the pixels on actually suppresses mode output, as generating negative weights is generally challenging in neuromorphic hardware. We can compare the fitted kernels to the activation maps shown in Fig. 5.4 d to intuit how a specific edge of an image may lead to the mode output being increased or decreased (with a notable limitation that no patterned illumination can produce less lasing than a fully dark

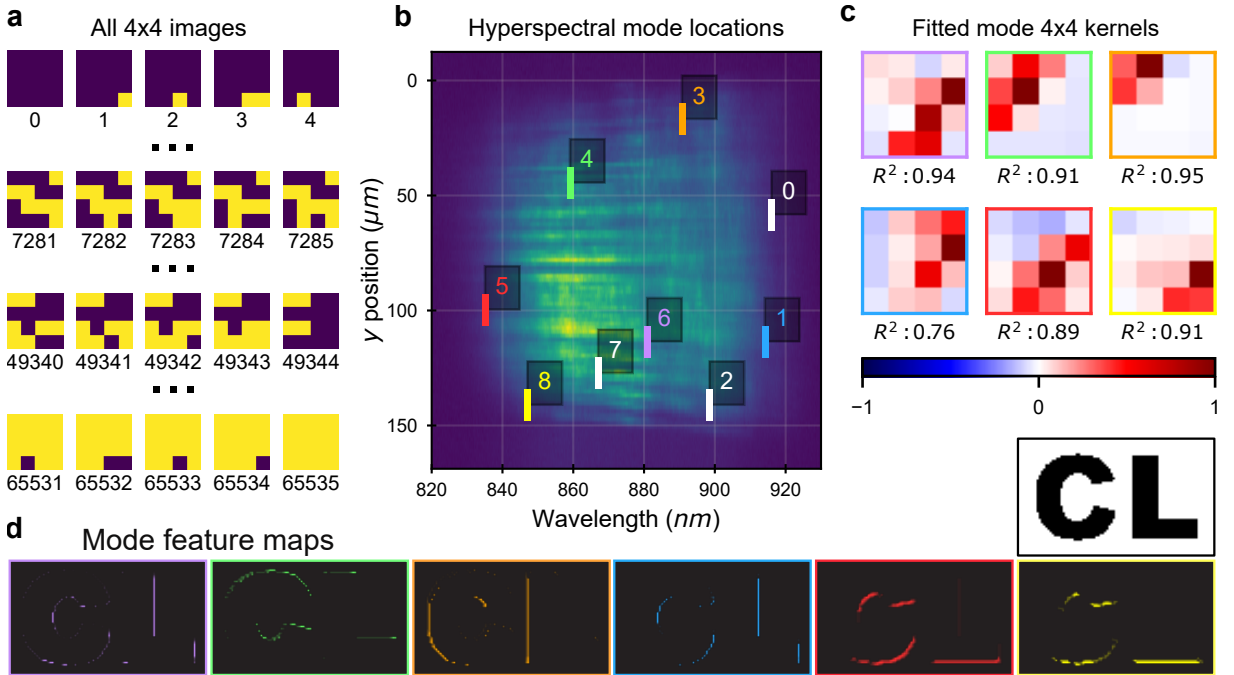


FIGURE 5.4: The 4x4 image library. **a** A selection of 4x4 black and white images, illustrating the method used to generate all possible variants. All of these images were projected onto a 150 μm lasing network, and the corresponding spectra were saved. **b** A full illumination spectrum from the library with selected ‘mode’ locations overlaid. These were chosen because they activate for relevant image features (edges for this investigation). **c** 4x4 kernels approximating 6 of the modes indicated in **b**, obtained by fitting to the experimental mode intensities. **d** Mode feature maps can be constructed for arbitrary black-and-white images by matching 4x4 raster scans of the image to the library. 6 mode maps of an image showing the letters ‘CL’ (for ‘Complex Lasing’) are shown, demonstrating that we capture all edges of the image. See Appendix Fig. B.5 c for activation maps of all modes.

image, where no output is seen).

An R^2 fitting score is shown under each kernel in Fig. 5.4 c. A software convolution is a linear process, and thus a similar fit applied to it would produce R^2 values equal to or slightly below 1 – this is demonstrated in the Appendix Fig. B.6, including for data with random noise added in. The kernels extracted from our experimental data have R^2 values of 0.628 to 0.946, indicating that aspects of the physical feature detection process cannot be fully captured by a linear description. As mentioned in Section 5.1.1, we believe this non-linearity in the feature detection process to be exciting for machine learning applications.

Since we can run any image through the library lookup process, we can also study how the selected modes react to features of our design. Fig. 5.5 shows an example of this, where I have generated an image of a rotating edge and processed it with the 4x4 library. The modes can be seen responding to the edge angle fairly smoothly, with each having a range of angles it responds to, and the full range of edge angles is covered by the selected modes. The ‘BW’ column corresponds to a binarised version of the mode activation maps, indicating what we can illuminate on the black-and-white DMD, which will be relevant for the two-layer *Machine Learning* (ML) scheme introduced in Section 5.3.

An image of white noise is also shown at the bottom of Fig. 5.5, aiming to test for false positive activation of the modes. As we have seen with the fitted kernels in Fig. 5.4 c, the response of the modes does not cleanly correspond to specific edges, and so some random activation is present, with the intensity of this varying by mode. The false positive ratio is comparatively not very high though, and does not appear to be a problem when processing real images.

The final column of Fig. 5.5 is a spectral mode, obtained as a region of the output spectrum, calculated by summing the spatial dimension of the measured hyperspectra (see Appendix Fig. B.5 b for the mode location). Spectral (instead of hyperspectral) modes would be a promising direction for the system, as a line array read speed is much faster than that of a 2D camera, but it appears that a lot of information is lost when we discard the spatial dimension, and I was only able to identify this single spectral mode.

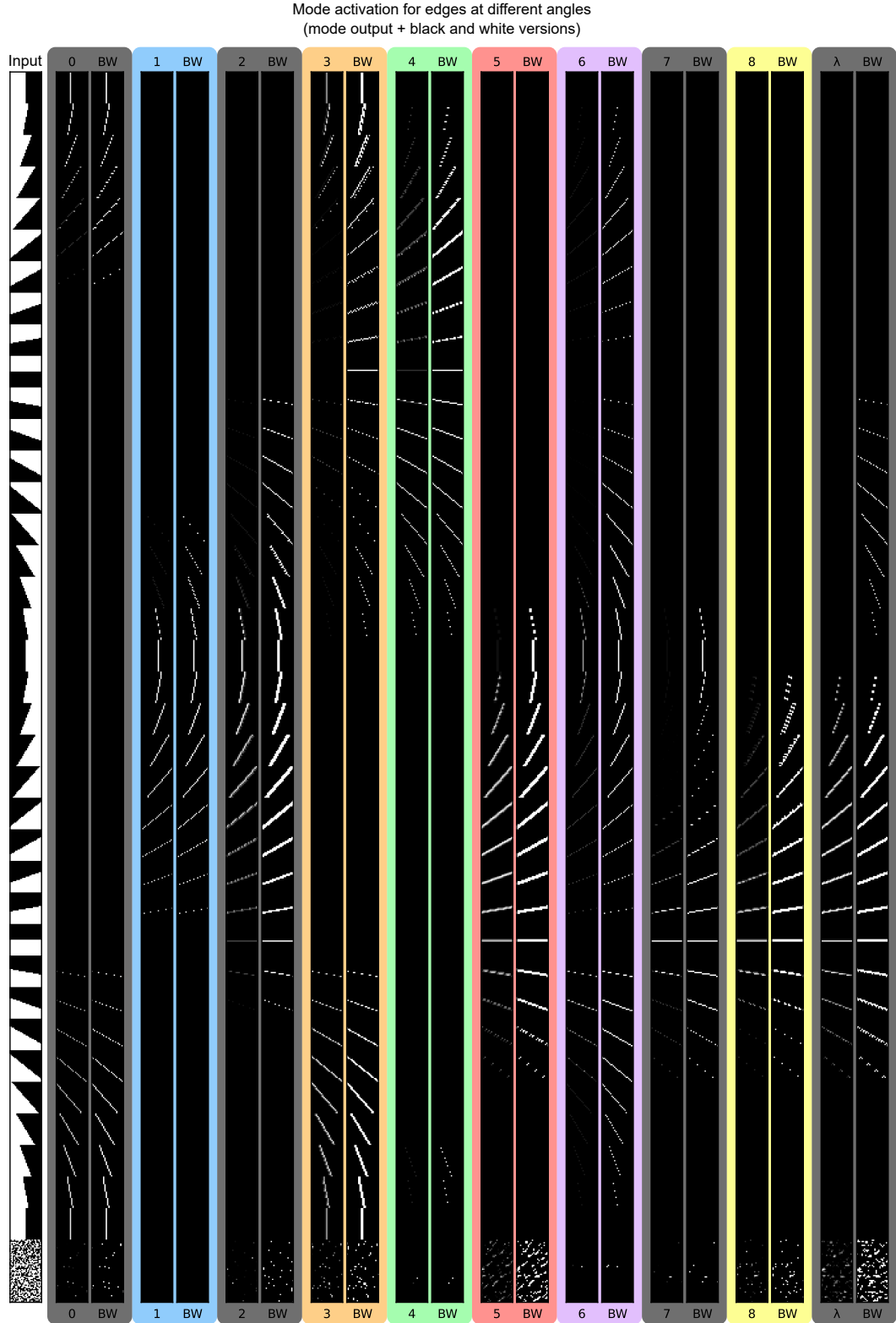


FIGURE 5.5: Experimental feature maps for modes 0-9 from the 4x4 library for input images (left-most column) of a range of rotated edges, and two white noise image inputs (bottom two rows). Each mode has two columns, an analogue-varying output produced by the mode-amplitude above the full illumination intensity (the mode ‘activation’, left column for each mode, labelled with mode number) and a binarised black and white version (right column for each mode, labelled ‘BW’).

5.2 Extreme learning machines – image classification

The feature detection performance of the network lasers tells us that the spatially-localised, mode-competing modes are excellent at processing spatial information in a non-linear way. The system is thus a prime candidate for more general image processing, and in this section I will explore how it performs as part of one of the neuromorphic architectures introduced in Section 1.2 – an extreme learning machine. In this application, the digital post-processing of the system output is limited to one linear layer, which would in general struggle to solve the problem on its own. The physical system thus has to do a large part of the computational heavy-lifting.

Inputting image data into the network is a simple extension of what we have already done for feature detection. Instead of projecting sections of images, we project an entire image onto the network, and observe the hyperspectral output. We started with the MNIST task, which is a handwritten digit recognition test commonly used in testing machine learning systems.^[174,175]

Fig. 5.6 explores the hyperspectra resulting from illuminating the first 10 000 digits from the MNIST dataset onto our network. The figures show average deviations of each class of hyperspectra from the overall mean hyperspectrum, thus highlighting the regions that are characteristically brighter or darker for each digit compared to the other classes. This shows that there are statistical differences between the network outputs corresponding to each class, and while these may be difficult to discern when looking at each spectrum separately, regression methods can be applied to reliably seek out these trends digitally.

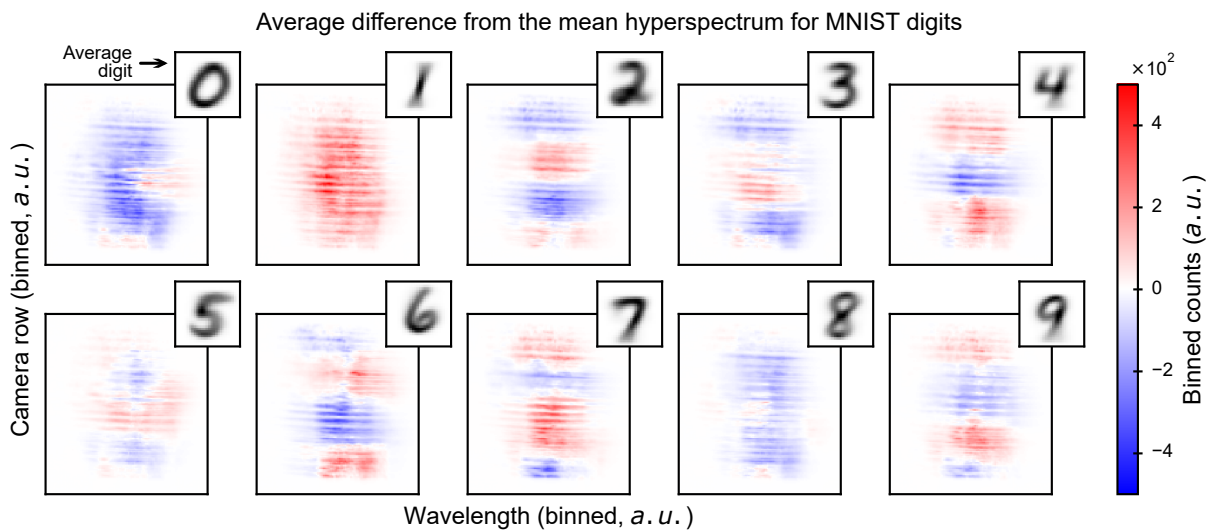


FIGURE 5.6: Average lasing hyperspectra for 10 thousand examples of the 10 classes of MNIST digits. Differences from the overall mean hyperspectrum are shown in red (higher than average) and blue (lower than average). Sub-plots in the top right of each spectrum show how the average member of each class looks like.

Linear regression is usually used to fit trends where the dependent variable is a real number. In classification problems, where the desired output is a categorical prediction, we often use *logistic regression*. This operates on a similar principle, but provides a model for converting the linear fit’s output to a probability value. We first establish a score s_k for each classification class k , which is a linear function of the input vector $\hat{x}$:

$$s_k(\hat{x}) = \alpha_k + \sum_i \beta_{ki} x_i, \quad (5.1)$$

with coefficients $\hat{\beta}_k$ and offsets α_k for each class k . Multiclass logistic regression then defines the probability that a sample $\hat{x}$ belongs to class k as:

$$P(y = k|\hat{x}) = \frac{\exp(s_k(\hat{x}))}{\sum_j \exp(s_j(\hat{x}))}, \quad (5.2)$$

where the denominator sum goes over all possible classes.^[176–178] We can see that summing the probabilities for all classes produces exactly $\sum_k P(y = k|\hat{x}) = 1$, as the numerators sum to the denominator, and the exponentials ensure that the semi-arbitrary value of equation (5.1) always maps to a positive value. The ‘weights’ $\hat{\beta}_k$ and α_k are established by minimising a cost function based on the probabilities of assigning the correct labels to a given set of training samples;^[177] this is performed by the ‘L-BFGS-B’ optimiser^[179] implemented by the Python package ‘scikit-learn’.^[180]

Notably, for a given $\hat{x}$ the denominator sum in equation (5.2) is the same for all classes k , and thus it is sufficient to compare the numerators to establish the class with the highest probability. Further, an exponential is a monotonic function, and so we can compare the scores from equation (5.1) directly, making classification through logistic regression a completely linear process equivalent to a single matrix multiplication per class.

Fig. 5.7 a illustrates how we apply logistic regression to utilise our network laser as an extreme learning machine. A picture of a 28x28 black and white MNIST digit is projected onto the network laser (colour inverted to increase the gain available to the laser) and a hyperspectrum is obtained. We apply two post-processing steps to the hyperspectrum to ensure optimal performance of the logistic regression fitting. First, the measured hyperspectra are scaled to remove the medians of the spectral regions and equalise their quartile ranges, which makes the fit converge more quickly. Second, we perform feature selection, discarding features that have low variance and score low on scikit-learn’s ‘f_classif’ feature-scoring function.^[180] The regression fit is run

multiple times with randomised training data sets (36 000 images from the exposed 40 000) in a hyperparameter study,^[181] where we vary the number of selected features and the amount of regularisation applied in the fitting.

In the end, the fitting of the linear regression gives us a ‘trained’ hyperspectrum classifier, which we can test on an unseen set of experimental hyperspectra. Fig. 5.7 **b** shows a confusion matrix resulting from this test, indicating a 96.03 % accuracy of our physical computing system for our best trial. When logistic regression is applied to the input images directly, it can reach accuracies up to 92.09 %, showing that the physical system contributes significantly to reducing the classification error rate (from 8 % to 4 %).

Fig. 5.7 **c** shows a selection of logistic regression weights obtained by fitting for 5 MNIST classes. Compared to the average hyperspectral trends shown in Fig. 5.6, these weight maps indicate small, specific spectral regions that make a specific spectrum more or less likely to be a result of projecting a digit from a specific class. Granular differences like these would be difficult to see by eye or in averaged data, but a computer is able to pick them out. Their presence shows the validity of our initial assumption – the non-linear and highly spatial dynamics of the lasing modes within our network lasers can map image inputs onto output modes in ways that we can pick up and use to provide useful information about the input.

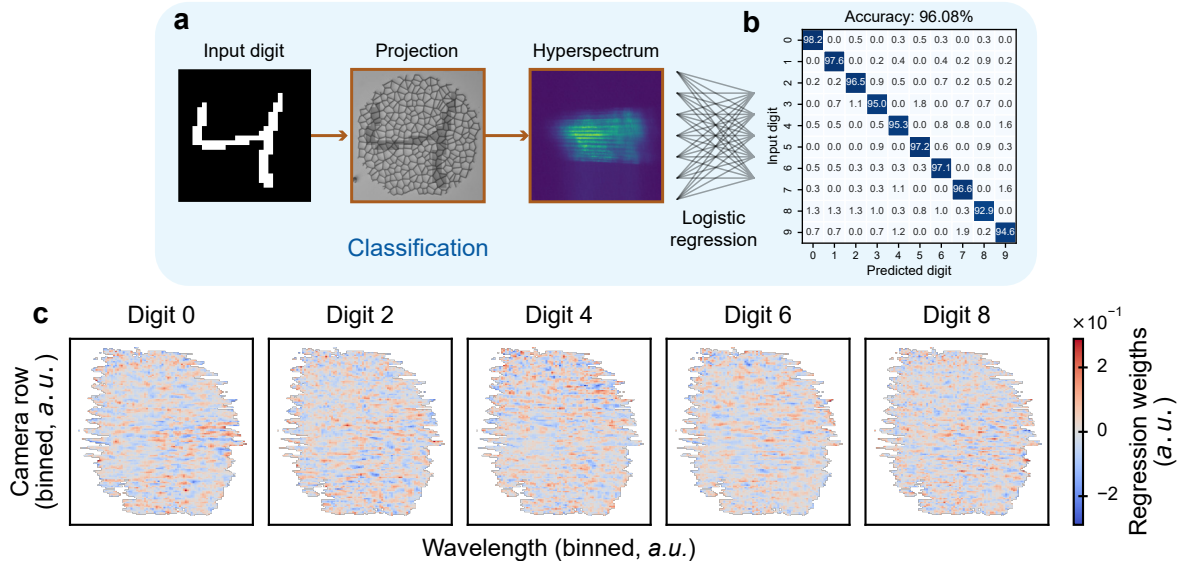


FIGURE 5.7: Single layer classification architecture. **a** The image to classify is projected onto the network laser, resulting in a hyperspectrum. With thousands of these spectra gathered, a logistic regression classifier (a single-layer, linear process) can be trained to predict the class of the image based on the spectrum (rather than the image itself). **b** A confusion matrix showing the result of this classification process for MNIST. **c** Logistic regression weights for several classes, showing how different hyperspectral regions are weighted when predicting whether the spectrum belongs to a class. Fully white regions were rejected during feature selection.

I will show single-layer classification results for tasks more complex than MNIST in Section 5.4, but first I shall introduce the second machine learning architecture we have developed.

5.3 Photonic Convolutional Neural Network

We have demonstrated that a network laser can perform feature detection and image classification tasks. In CNNs these tasks are done in sequence: first, convolutions are used to tease out the features of the input image, and then a *multilayer perceptron* (MLP) step uses these features to classify the image. We can employ our physical analogues of these two steps to construct a Photonic CNN for image classification. Fig. 5.8 illustrates this architecture: an initial feature detection layer (Fig. 5.8 a), then a second layer where the detected feature maps are illuminated on the network as a single whole image (Fig. 5.8 b). A hyperspectrum from the original input image is also included in the result pool, and a logistic regression step is applied to the spectral output of the second layer to generate classification results (Fig. 5.8 c). This is repeated for all 40 000 images we use for this test. The classification accuracy is shown to improve from the single layer 96.03 % to the two layer 98.05 % for our best trial, further reducing the classification error rate from 4 % to 2 %.

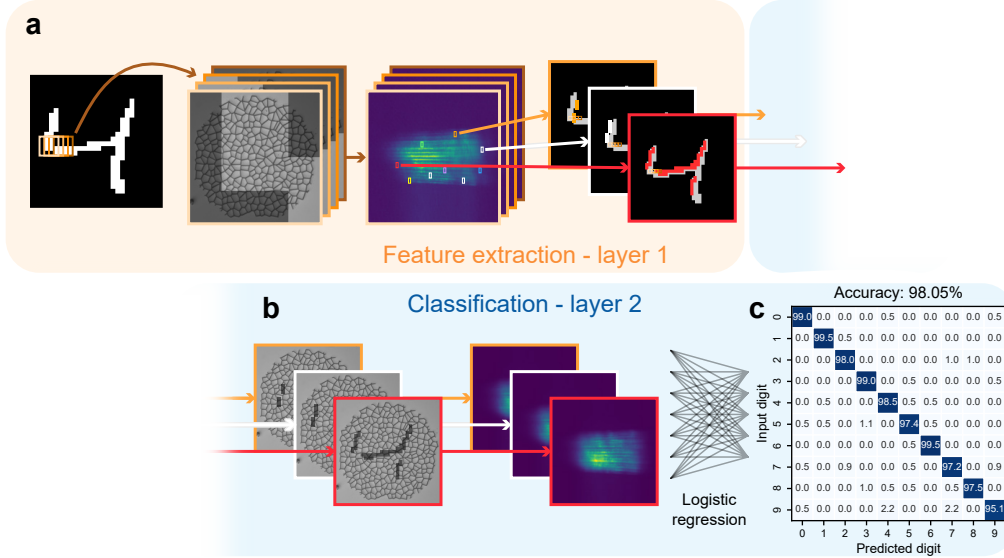


FIGURE 5.8: Two layer architecture – a Photonic Convolutional Neural Network. **a** The images to be classified are raster-scanned to extract features like edges (this is currently done using the 4x4 image library to accelerate the process). **b** The extracted edge maps (plus the original image) are projected onto the network to produce corresponding hyperspectra. The logistic regression is now trained on this expanded set of spectra. **c** Using the two-layer approach shows an increase in MNIST accuracy.

As discussed in Section 5.1.2, acquiring data for all the images required by a raster scan

takes a comparatively long time in our current experimental conditions. Having to do over 500 additional exposures for each of the 40 000 digit images in this dataset would not be practical. Fortunately, we can use the 4x4 image library from Section 5.1.2 to generate equivalent feature maps using real experimental data (example maps shown in Fig. 5.8 a, and maps for all modes and classes shown in Appendix Fig. B.7). This allowed us to accelerate the measurements significantly at this stage, but will hopefully not be needed in the future as we accelerate the experimental processes.

The Photonic CNN two layer scheme demonstrates the versatility of our physical system for machine learning tasks – both of the layers involve a very different neuromorphic process running on the same physical hardware, and the fact that we gain additional information from the feature map projection hyperspectra indicates that the system is suitable for processing a broad variety of input images.

5.4 Harder tasks and few-shot regimes

MNIST can be considered an easy machine learning task – it is usually used to test basic classification systems, and is fairly linearly separable, as logistic regression alone achieves 92.10 % accuracy in our tests. To fully understand the performance of our physical neuromorphic system, we need to test it on tasks where the final linear layer contributes less to the classification. This section explores two tasks that are more challenging, and compares our performance directly to logistic regression and several software ML benchmarks.

Fashion MNIST^[182] is a task developed to be a drop-in, more difficult replacement for MNIST. It consists of 28x28 greyscale images of clothing belonging to 10 classes. Several examples of the images are shown in Fig. 5.9 a, along with features detected using the experimental 4x4 image library. Compared to MNIST, the complexity of the images is much higher, with small differences sometimes determining the difference between a t-shirt, shirt, and pullover. While our current edge detection scheme is limited to black and white images, it manages to pull a good amount of detail out of the clothing items (see Appendix Fig. B.8 for edge maps for images from all 10 classes).

The classification performance of our network lasers is shown in Fig. 5.9 b, along with the 10 class labels. For the single layer scheme we project the images in approximate greyscale, using ordered blue noise dithering^[183–185] to represent the images on the black and white pixels of the DMD – in video applications, the DMD pixels are repeatedly flicked on and off to achieve grey

levels through persistence of vision, but this strategy cannot work with femtosecond pump pulses. The ratio of image resolution to DMD resolution allows us 145 levels of grey per data pixel with the dithering method, and the average representation accuracy improves over larger areas. The single layer test yields an accuracy of 85.18% in the best trial, and the two layer process gives 87.85%. This is lower than the accuracies for MNIST, but fairly high taking into account the increased difficulty of the task – logistic regression alone achieves 82.29% for Fashion MNIST, and the benchmark for human performance (given no fashion expertise) is 83.5%.^[182]

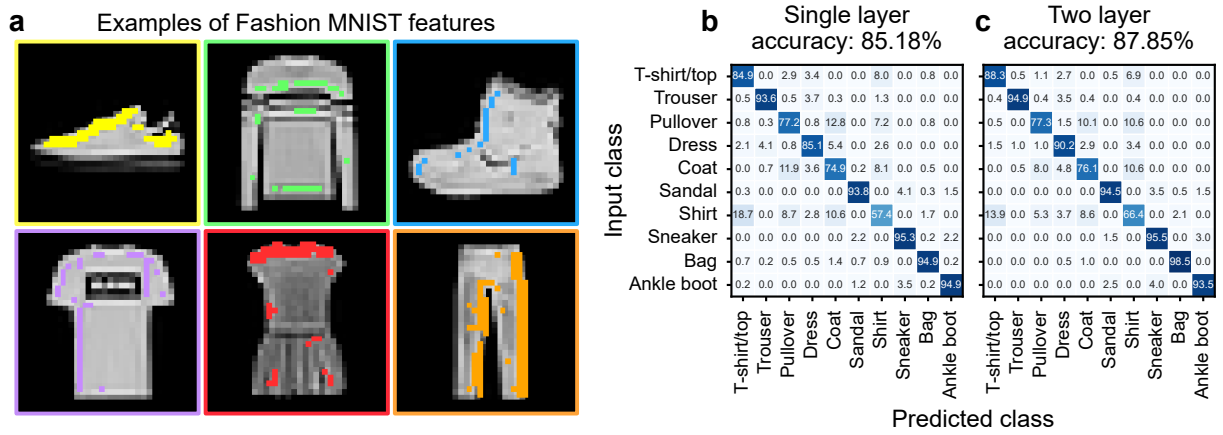


FIGURE 5.9: Classification results for Fashion MNIST, a dataset consisting of photos of 10 types of clothing. **a** Examples of the edge maps extracted from six Fashion MNIST images. **b** Accuracy for the single-layer approach. **c** Improved accuracy from the two-layer approach.

There is much interest in applying machine learning to medical tasks, whether for analysing patient data, diagnosis and segmentation in medical imaging,^[186,187] or some proposed applications where computer methods augment the human decision making process.^[188] ML models promise accuracies comparable to experts but at much higher speeds, or from statistical trends in imagery that would be difficult to spot by eye.^[189] A significant challenge for this research area is the availability of high quality, large datasets. Typical ML studies employ tens or hundreds of thousands of images to train their models, but typical biomedical datasets need to be gathered by hand and with the patients' consent, usually resulting in a much smaller sample size.^[186,190]

To test our performance within these constraints, and for images more complex than MNIST or Fashion MNIST, we have tested the network's classification performance on the BreakHis 400X task – a database of 1 670 breast cancer histopathology images taken at 400x magnification adapted by Ref. [191] from Ref. [192], and further augmented through rotations to 2 246 images for some tests here (in order to balance the number of samples in the two classes). Several example images are shown in Fig. 5.10 **a**, illustrating the visual complexity of the data, as well

as how difficult it is to tell apart the two classes ('benign' and 'malignant', indicating the lack or presence of breast cancer cells) by eye. I have further rescaled the images to be 190x190 pixels, to better match our DMD resolution and the spatial sensitivity of the network.

This is a colour dataset, and so I have processed it by splitting it into its three **RGB** channels. As first shown in Fig. 5.3, this can allow us to detect features specific to each channel, and the feature maps overlaid in Fig. 5.10 **a** show that this is the case for the histopathology images too. For example, the selected 4x4 library mode highlights voids in the green channel, and finer cell/nucleus details in the red and blue channels. The images themselves are also illuminated onto the network as three separate illuminations, one per **RGB** channel, and the resulting hyperspectra are combined for the logistic regression step. Blue noise dithering is again used to represent greyscale images, with 4 **DMD** pixels per data pixels. While this leads to an average absolute brightness error of 12.5 % for each data pixel (as the 256 original grey levels are quantised to only 5), the nature of dithering is that these error average out over larger areas.

Our classification performance for the two neuromorphic schemes is shown in Fig. 5.10 **b**. The single layer classification accuracy is 90.16 % in the best trial, with the two layer process reaching 92.94 %, the correct classification of the malignant class improving in particular. The logistic regression performance when trained directly on the images is only 70 % to 74 %, showing that the physical systems plays an increased role in this task. Appendix Fig. B.9 shows the

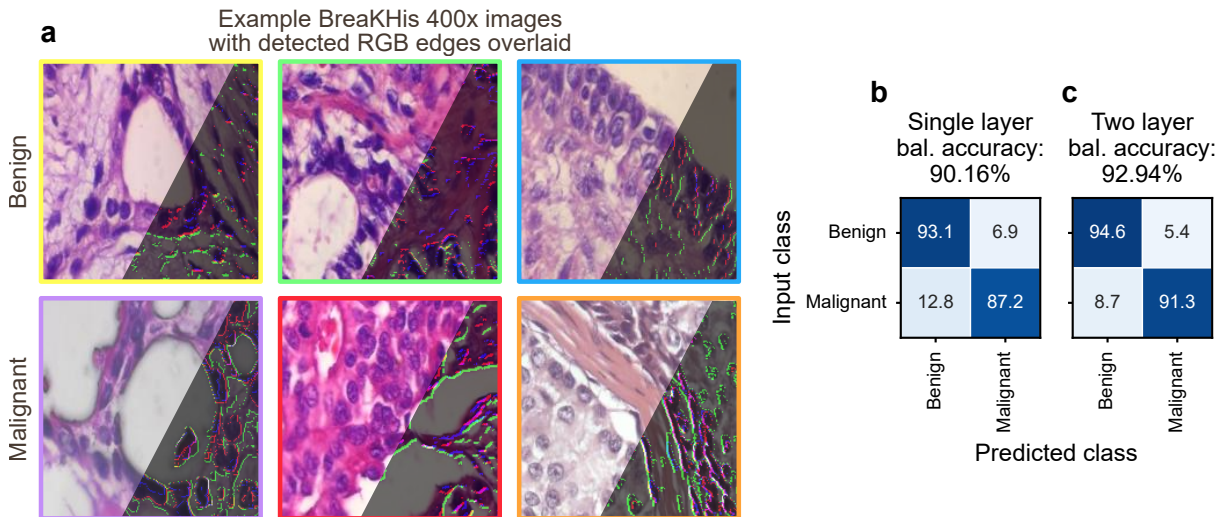


FIGURE 5.10: Classification results for BreakHis 400X, a dataset of microscopic biopsy images for detecting breast cancer. **a** Examples of benign and malignant images from the dataset, with edge maps from several modes of the 4x4 library overlaid. The images within each class can differ from each other significantly, and the difference between classes is not immediately obvious. The dataset is provided in colour, and so we extract separate mode maps for the three colour channels, and expose each channel separately. **b** Single layer physical system accuracy. **c** Improved double layer accuracy.

regression weights for this task, indicating that both the original image and mode illuminations across all three colour channels contribute to the final prediction.

Comparing our performance to logistic regression gives us an idea of how much of the system performance can be reproduced by the final linear layer alone, but we are also interested in how the physical implementation compares with better digital ML models. Fig. 5.11 shows a comparison of classification accuracy for our three classification tasks vs. training dataset length for a range of software neural networks and our photonic network. The digital models shown are: logistic regression, a *multilayer perceptron* (MLP), and two software CNNs: one following the LeNet5 architecture,^[175] and a larger CNN with more filters and neurons (see Methods in Ref. [67] for details, and the legend in Fig. 5.11 for the number of trainable parameters). Training dataset length for MNIST and Fashion MNIST is varied from 125 to 36,000 images, with a fixed 4000 image test set.

Our system outperforms logistic regression over the full range of training lengths, proving that the physical layer contributes significantly to the ML performance. For MNIST and Fashion MNIST (Fig. 5.11 a and b) we achieve accuracies comparable to the digital systems, especially with the boosted performance of the two layer system. Notably, the photonic network achieves higher accuracy than software benchmarks at shorter training set lengths – this is highlighted in the zoomed-in plots. We call this the ‘few-shot’ regime, where training data is scarce. As we see

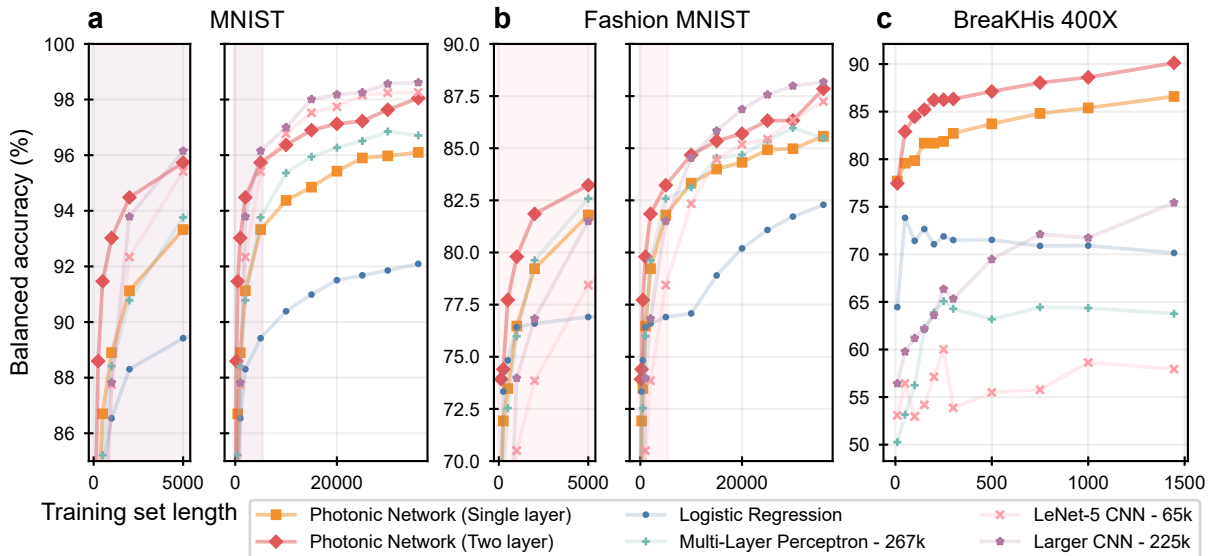


FIGURE 5.11: Accuracy vs training length for our physical system and fully digital solutions. Results shown are for **a** MNIST, **b** Fashion MNIST, and **c** BreakHis 400X. In **a** and **b**, the left column shows a zoomed-in view of the ‘few-shot’ regime, where limited training data is available – the photonic network outperforms digital benchmarks when working with a small number of images. The legend lists total trainable parameter numbers for the three neural network-based digital benchmarks.

with BreaKHis 400X, medical datasets often require good few-shot performance, as very little data is available. In fact, for this small dataset we outperform the digital benchmarks tested for all training lengths available with the 1 670 images, as shown in Fig. 5.11 c.

In addition to medical data, few-shot learning is also critical for ‘edge computing’ applications, where remotely deployed hardware performs computation directly at the point of sensing. This tends to require learning from limited data, while adapting on-the-fly to changing environmental conditions, all without access to high-performance compute clusters. Digital neural networks generally struggle when there is a very high degree of mismatch between the amount of data available for training and the number of weights we are attempting to train – too many parameters can result in overfitting, where the model uses the abundant weights to fit the training set exactly, but performs badly on the test set due to poor generalisation.^[193] This explains the poor performance of the digital benchmarks on BreaKHis 400X seen in Fig. 5.11 c, where even logistic regression outperforms the over-parametrised larger models if very few training examples are available. One of the ways this problem is dealt with in machine learning is ‘augmentation’, where large quantities of additional data are generated by rotating, scaling, and otherwise distorting the original images.^[190,193] This technique has been successfully applied to medical tasks like BreaKHis 400X,^[194] but requires significant additional computation, especially for larger models. Our system’s ability to perform well on short training sets directly (without augmentation) is thus a potential benefit, pointing towards real-life applications where our scheme could be useful.

5.5 Physical basis for neuromorphic performance

We have shown that our network lasers can perform quite impressively compared to digital benchmarks, especially in the few-shot regime. While the extreme learning machine approach to neuromorphic computing allows us to treat the system as an effective black box, understanding the physical reasons for our good performance remains desirable. This is especially true if we want to optimise the system in the future – understanding its workings would allow us to focus on specific avenues of tuning, redesigning the topology or introducing biasing to better align the physics of the system to the task at hand.

Neuromorphic systems often struggle to implement non-linearity and negative weights. We address the linearity challenge both with the intrinsic non-linearity of a laser, and the non-linear coupling between our many lasing modes. Our feature detection application (introduced in Fig. 5.1) proves the strong mode competition in our system. Through coupling and interactions

between the lasing modes, we achieve highly non-linear behaviour (as first predicted theoretically in Fig. 4.9). This gives us the high sensitivity to patterns in the spatial illumination, shown in Fig. 4.11 a, as well as to the intensity of the pump, as shown in Fig. 4.8. Together, these sensitivities map complex, greyscale images into unique hyperspectra, sensitive even to spatially small changes (like the BreaKHis 400X edge maps shown Fig. 5.10 a).

We also show that the mode competition enables modes to suppress each other, resulting in certain modes responding negatively to positive inputs, as seen in the statistical analysis of the 4x4 image library in Fig. 5.4 c. The non-linearity and negative weights of our networks could thus be a part of our neuromorphic performance advantage.

Other neuromorphic researchers also find that the number of parameters needed to tune the output of their systems is much reduced compared to the full digital classifiers – Ref. [45] reports a number of digital parameters up to 97% smaller for their physical system than the equivalent digital benchmarks (Appendix Fig. B.9 shows a preliminary study of how this looks for our system). One way to interpret this is that most of the neural network weights necessary to solve the problem have been ‘moved’ to the fixed physical data processor. Intuitively, this approach works best when the physical system can thus ‘encode’ a large number of weights and complex relationships between them. A recent theoretical study proposes that a system that represents its inputs in a very high-dimensional space can be an explanation for strong few-shot performance.^[195] We thus believe that the highly complex modal landscape of our system (studied in Fig. 4.3) could be a part of how the lasing network manages to make the tested ML tasks linearly separable – it is a high-dimensional transformation that ‘understands’ the spatial relationships within data with no training required, and is thus intrinsically suited to vision tasks. With most of the processing happening in the fixed physical system, the few-shot performance can be understood more intuitively as well. The laser itself is independent of the amount of data available, so should be more resilient to reducing the training set, which impacts only the linear trainable classifying layer.

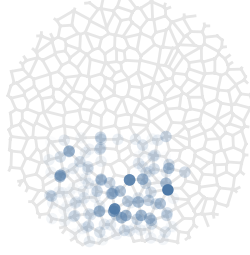
Neuromorphic computing also studies how brain-inspired systems can be used to accelerate ML, and so we can compare our system to aspects of real neurons. In the human retina, feature detection is performed by retinal ganglion cells, the final output neurons in the retinal circuitry.^[196] These cells utilise lateral inhibition to non-linearly enhance image features such as edges.^[197–199] When a ganglion cell is activated, it sends excitatory signals to the brain while simultaneously sending inhibitory neurotransmitter signals to adjacent ganglion cells, decreasing their firing activity. Different subsets of ganglion cells respond preferentially to different angles

and orientations of edges, allowing the retina to process complex images while detecting multiple features in parallel, with non-linear coupling between the detection processes.^[196,200] The close match between the feature detection mechanism of ganglion cells and network lasers makes for an interesting parallel, showing that our vision mechanics are in some ways similar to that of the brain – we thus refer to our approach to neuromorphic computing as ‘retinomorphic’.

Recent research also shows that heterogeneity in neuronal systems can lead to better and more robust machine learning performance, especially on complex tasks.^[201] The ‘neurons’ (modes) of our laser are all different from each other and coupled in different ways, which we think could confer a similar advantage, especially given that our performance compared to digital benchmarks also improves for more complex tasks like BreaKHis 400X.

Conclusions

In this chapter we have demonstrated that semiconductor network lasers are a fitting platform for neuromorphic computing. They are capable of feature detection and image classification, and combining the two capabilities leads to even stronger performance. The feature detection gives us unique insight into the physics of the lasing network, illustrating a practical application of mode-competition and its suppressive mode coupling. Image classification shows that the complex mode dynamics enable the network to process data spatially, in a way that is especially valuable when limited training data is available. While I do outline some potential physical bases for the network’s performance, exploring how the neuromorphic application links to the underlying physics is an interesting topic of study with much potential.



Chapter 6

Conclusions

6.1 Summary

In this thesis I have demonstrated the fabrication, characterisation, and neuromorphic application of semiconductor network lasers. The system has proven to be a very promising platform for computing, and the stability and flexibility of the semiconductor platform opens many doors in terms of optimisation and exploration compared to previous implementations of network lasers. This section presents an overview of the main outcomes for each of the thesis chapters, which all contributed to developing this novel platform.

Chapter 2 showed the fabrication details of our III-V integration process, *Direct Wafer Bonding* (DWB). A number of existing low- to mid-temperature etch recipes were explored for processing bonded layers of InP. The cyclic CH_4 recipe was shown to be versatile and reliable. A continuous CH_4 recipe was proposed to address some shortcomings of the cyclic approach, with the main result being a *Design of Experiment* (DoE) study of the impact of etch parameters on the process outcomes. The versatility of this approach was demonstrated by tuning the continuous CH_4 etching process to work with *quantum well* (QW) layers epitaxially embedded in the etched InP layer.

In Chapter 3, the etched devices were put to the test. Lasing from *Whispering Gallery Modes* (WGMs) in etched microdisks was demonstrated, and the lasing threshold explored as a metric for etched cavity quality. Oxygen cleaning of the devices was shown to systematically decrease lasing thresholds. For QW lasing disks, H_3PO_4 treatment and Al_2O_3 passivation coating were both shown to improve device output and reduce degradation under longer exposure, with H_3PO_4 treatment additionally lowering the device threshold. Overall, both continuous and cyclic

CH₄ etches were shown to produce well-performing nanolasers, with the QW disks even displaying state-of-the-art *continuous wave* (CW) performance. The etching studies in Chapter 2 and the performance verification in Chapter 3 validate our chosen path for implementing high-quality semiconductor nanolasers.

Chapter 4 showed that we can etch working network lasers out of InP layers, and explored their physics. The mode landscape of the laser was shown to be extremely complex, with mode shapes obtained both experimentally and through simulations. Mode competition was shown to be a major driver of system non-linearity, with some modes lasing more strongly under partial illumination. The transport and losses within the network were studied, showing how coupling between waveguides depends on the waveguide width, and that node losses are a significant factor to consider – unamplified light would be lost after traversing only a couple of network nodes. Long-distance coherence is nevertheless enabled by stimulated emission, with modes spanning large areas of the network, and distant parts of the system depending on each other. The losses still drive the system behaviour – this was experimentally demonstrated by disrupting the coupling between network halves with a low-gain region. More broadly, the system was shown to be sensitive to the gain landscapes obtained by patterned pumping, with both losses and mode competition shaping the highly variable output spectra. We also showed that the networks are resilient to degradation when pumped for long times, and that many network topologies exhibit lasing, so there is much room for explorations beyond our currently preferred Voronoi-based design.

Chapter 5 applied the physics of network lasers to a number of neuromorphic computing tasks. Feature detection was shown first, where different modes in the laser’s hyperspectrum ‘activate’ when specific image features are raster-scanned across the system. This can detect a number of features in parallel, and the flexibility of our approach was demonstrated through edge detection at multiple raster window sizes and across RGB image channels. A statistical study of the modal response was presented, where equivalent convolution kernels were fitted to hyperspectra from all possible 4x4 black and white images. This library of hyperspectra also allowed us to reconstruct experimental edge maps for arbitrary images, speeding up further analysis and classification experiments. Finally, we demonstrated the application of a network laser as an extreme learning machine for image classification. This was shown to work well as a single layer scheme, with the performance improved further by including a feature detection step in the process, forming a double layer ‘Photonic CNN’. We demonstrated classification performance of 98.05 % on MNIST (digits), 87.85 % on Fashion MNIST (clothing), and 92.94 %

on BreaKHis 400X (breast cancer histopathology), achieving particularly good performance in the few-shot regime.

6.2 Outlook and future work

We believe this work lays a strong foundation for the use of semiconductor network lasers as a neuromorphic computing platform. There are many possible directions for investigations that build on it, either for studying random lasing in general, or for optimising computation. In this section I will discuss some general challenges that can be addressed in future work, and a few projects that I believe would be exciting evolutions of this effort.

6.2.1 Energy use and speed

The goal of neuromorphic computing is accelerating machine learning tasks beyond the abilities of modern GPUs. The relatively recent emergence of the field does mean that current implementations do not live up to that ambition yet, focusing instead on establishing the physics and technology, and potentially solving problems in specific niches. Our system is also in the fundamental research phase, but it is interesting to compare it to classical computing nevertheless, and chart a path forward for future improvements on this front.

The energy use and time scale of the underlying physics in the network laser are both very small. The energy delivered to the sample during our machine learning tests is in the range of 35 nJ /pulse, and we expect the lasing dynamics to happen on the timescale of 10 ps to 100 ps. Experimental constraints mean that the test setup as a whole does not operate in this regime. First, our pump lasers are powerful femtosecond sources, consuming much more energy than is delivered to the chip. We have been testing smaller picosecond pulse diode laser systems, potentially bringing the pump’s power consumption down from 600 W to ~ 15 W – while the networks do lase, as shown in Appendix Fig. B.2, increased degradation from the longer pulses is currently an issue. Nevertheless, with further fabrication improvements, like increasing heat dissipation to the substrate or minimising side-wall recombination, this is a realistic path towards miniaturisation and energy efficiency. Timing-wise our main bottleneck is measurement time, with cameras having to integrate over multiple excitation pulses (on the order of 80 to 300 in most of our testing), and readout speeds being limited. I have made significant strides here, with software and hardware improvements bringing our main setup from operating below 1 Hz to ~ 7.5 Hz, and our new setup, built by W. K. Ng and programmed by me specifically to accelerate

these experiments, has reached 100 Hz. Further improvements are possible – the limit is almost entirely how quickly we can grab frames from a camera with a high enough signal to noise ratio.

Currently the most exciting angle to look at our potential as an [AI](#) accelerator is the few-shot performance of our system, and the low training costs. While pushing the data through the system is not entirely optimised, training a single logistic regression layer is much less computationally intensive than training a larger neural network. For datasets with limited examples available, like BreakHis 400X, classical systems also usually rely on ‘augmentation’, where large quantities of additional data are generated to aid the training. Our physics appears to enable us to work without this augmentation, presenting a potential advantage as classical systems spend more time and energy processing the additional data.

6.2.2 Material and fabrication

Further material and fabrication explorations could yield better-performing network lasers. I was unable to thoroughly test continuous CH_4 etched InP network lasers, as we have focused on utilising the cyclic-etched chip which I delivered earlier in the project, but preliminary measurements indicate that the continuous-etched networks perform comparably to cyclic-etched counterparts. It is possible that the smoother continuous side walls would lead to less device degradation, enabling picosecond pumping for example. Similarly, [QW](#)-based network lasers are a promising avenue to explore, with narrower lasing peaks potentially providing a more direct way to map hyperspectra to analogue outputs. While the devices I have fabricated do not lase due to insufficient gain, we are working with collaborators who can fabricate chips with more quantum wells and well-tuned etch recipes. If these do lase, there is much potential for new neuromorphic approaches, and applications of network lasers closer to telecommunication wavelengths.

An exciting ambition for the long-term fabrication of network laser devices would be to integrate on-chip pump light sources with the network, aiding miniaturisation and lowering energy use.

6.2.3 Tuning and network topology

The presented network systems work well with very limited tuning – a network was selected with a rough parameter sweep of diameter and density, without a detailed study of how these parameters impact performance metrics, and the machine learning tasks we study appear to

work well on a variety of network shapes and topologies. There are some limitations still. In simulations, up to 40 hyperspectral modes exhibit feature detection for smaller networks, but we were able to observe up to 9 experimentally. In classification tasks, we perform very well in the few-shot regime, and match digital benchmarks of reasonable size when more data is available. More complex datasets like CIFAR-10^[202] (vehicles and animals) remain a challenge. Pre-fabrication and post-fabrication optimisation could let us extend the good performance beyond the few-shot regime, matching or beating larger models too.

Changing the network topology would let us tune the networks pre-fabrication. Our intuition is that increasing the amount of disorder in the network could lead to better performance, and there are parameter sweeps on our new samples which are designed to test this (illustrated in Fig. 6.1 a). Establishing more standardised performance metrics, and correlating them to the networks’ graph-theoretical properties, could give us an understanding of what topologies work best and how to optimise them. We could then design networks specifically optimised for the tasks we are interested in handling.

Post-fabrication tuning would aim to adapt the already-made networks to match their mode landscapes to specific tasks. This can be achieved by introducing additional weights into the system, which are optimised to improve some metric, but do not change with the input data – this principle is illustrated in Fig. 6.1 b, and has been shown to work for neuromorphic systems, for example in Ref. [45] and Ref. [51]. The high sensitivity of network lasers to small variations in the material properties^[63] and illumination^[65] indicate that they would be highly susceptible to this tuning approach.

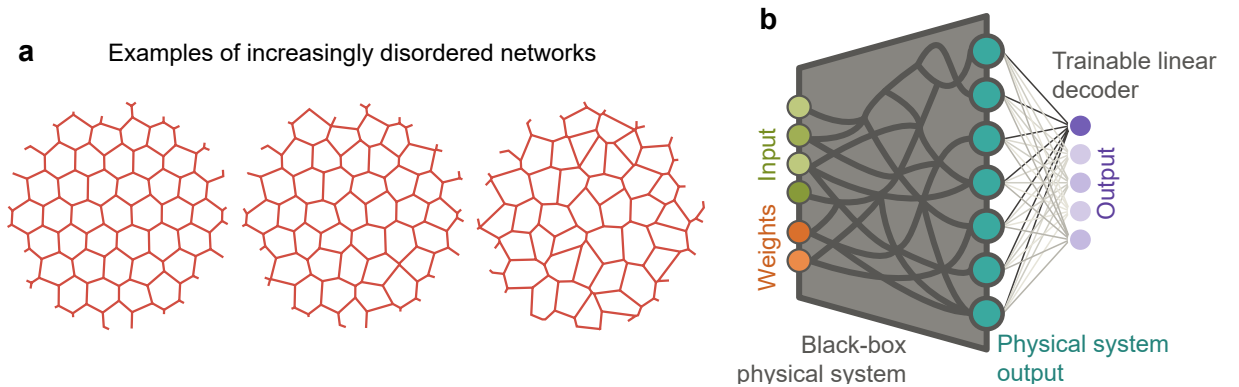


FIGURE 6.1: Pre- and post-fabrication tuning opportunities. **a** Networks with different graph-theoretical properties can be fabricated, for example varying the amount of disorder. **b** Tuning an extreme learning machine. A set of ‘weights’ is provided with the input and optimised to improve output quality. Once good weights are found, they are fixed and do not vary with data input. If the physical system can be tuned well enough, it may also be possible to remove the linear decoder layer and use the system’s output directly.^[45]

6.2.4 Segmentation and other novel tasks

There are many task types in the field of machine learning. We have demonstrated that our system can perform well on three image classification tasks: MNIST (digits), Fashion MNIST (clothing), and BreaKHis 400X (breast cancer histopathology). I have also tested two other tasks in this category. HAM10000^[203] is a heavily-imbalanced skin lesion classification dataset, with only a small number of examples included for some of the seven classes. Our single layer classifier reaches 53.7 % balanced accuracy, which is a good start, especially compared to digital benchmarks which struggle without augmentation – EfficientNet-B0,^[204,205] a large classifier, only gets around 45 % accuracy in our tests. CIFAR-10^[202] consists of 32x32 colour images of vehicles and animals divided into 10 classes. Here, we get up to 40.9 % accuracy, showing that there are limits to our un-tuned performance.

Tasks other than image classification are also possible. Our team has previously looked at non-linear waveform transformation.^[206] Feature detection could be extended to kernels more complex than image edges, and potentially greyscale data as well. Non-image tasks could be achieved by converting numbers into projection brightness values, potentially encoding time-series data too.

A task that we have exciting preliminary results for is segmentation, where the system tries to draw a boundary around a specific projected image feature. We have demonstrated that this is possible with HAM10000 data, as illustrated in Fig. 6.2. The images are projected onto the network, and a ridge regression layer is used – we set the output layer size to be the same as that

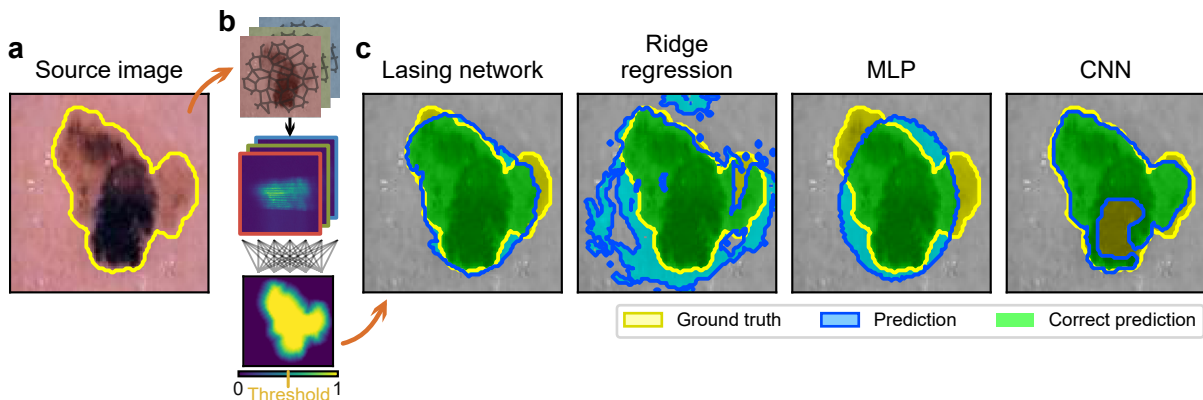
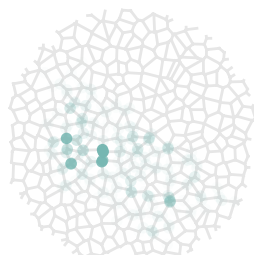


FIGURE 6.2: Network lasers for image segmentation. **a** An example skin lesion image from the HAM10000 dataset, with the lesion outline drawn in yellow. **b** The experimental segmentation process. The RGB channels of the image are projected onto the laser, obtaining hyperspectra. A ridge regression is trained on the hyperspectra to match segmentation masks in the training set. Confidence maps are computed for each hyperspectrum, and a threshold can be applied to obtain a binary mask. **c** Segmentation results from the photonic network laser and three digital benchmarks.

of the input image. We then train the linear layer to reproduce masks provided by Ref. [207,208], obtaining confidence maps of where the system ‘thinks’ the skin lesion resides (Fig. 6.2 **b**). A threshold can be applied to these, giving the prediction shown in Fig. 6.2 **c**. We perform better than digital benchmarks across a couple of different metrics, with detailed results being prepared for an upcoming publication, showing that there is much potential for applying network lasers across a range of neuromorphic workloads.



Appendix A

Experiment automation – puzzlepiece

Many of the results presented in this thesis rely on complex measurements: automatic device array scans and alignment for disk lasers, or illuminating hundreds of thousands of images onto the network lasers. Creating sustainable and flexible software for automating these processes is a problem that often arises in laboratory environments, and I have set out to attempt to provide a solution to it during my PhD work.

The result of this work is ‘puzzlepiece’, a Python framework for rapidly building modular experiment *Graphical User Interfaces* (GUIs), which is now open-source, and available on the Python Package Index. I will provide a brief overview of my design choices and the framework’s functionality, but the full documentation is available here:

<https://puzzlepiece.readthedocs.org>

First, it was important to me that the experiments can be controlled as fully as possible from the GUI. Many users of experimental setups prefer to operate on visual interfaces rather than code, and generally during alignment and investigative work, being able to click a couple of buttons is faster and easier. Creating GUIs generally is a rather complex and repetitive process though, using much boilerplate code. The first goal of puzzlepiece was thus to automate the generation of components like number input boxes, checkboxes, and buttons. These are then placed automatically on a standard grid, so that the software is standardised, and the programmer does not have to think of the layout.

The second principle of puzzlepiece is that accessing all options should happen through a unified *Application Programming Interface* (API). Equipment manufacturers usually provide their own API, which allows software to talk to the device. Much of the time, this API does

not follow a specific standard, making a case where one needs to interface with multiple devices quite complex – the spectrometer may use a C-like interface, the camera have a Python module, and the temperature controller communicate over serial. With puzzlepiece, a programmer can create a uniform abstraction layer over these by creating setter and getter function for device parameters. That way, once the hardware communication layer is implemented, all parameters across all devices can be changed and obtained using the same `set_value` and `get_value` methods. This process also creates GUI components for the parameters, and so the device can be easily interfaced through code and through the GUI. An example of a hybrid setup like this is shown in Fig. A.1, where a Jupyter Notebook works side-by-side with the GUI, enabling a powerful workflow where alignment and inspection are done interactively, and more complex measurements programmatically. The possibility of running GUI event loops within IPython kernels was first brought to my attention by Thomas Karg. I have used this option heavily throughout my PhD work, and submitted patches to the IPython project to make the experience more seamless.

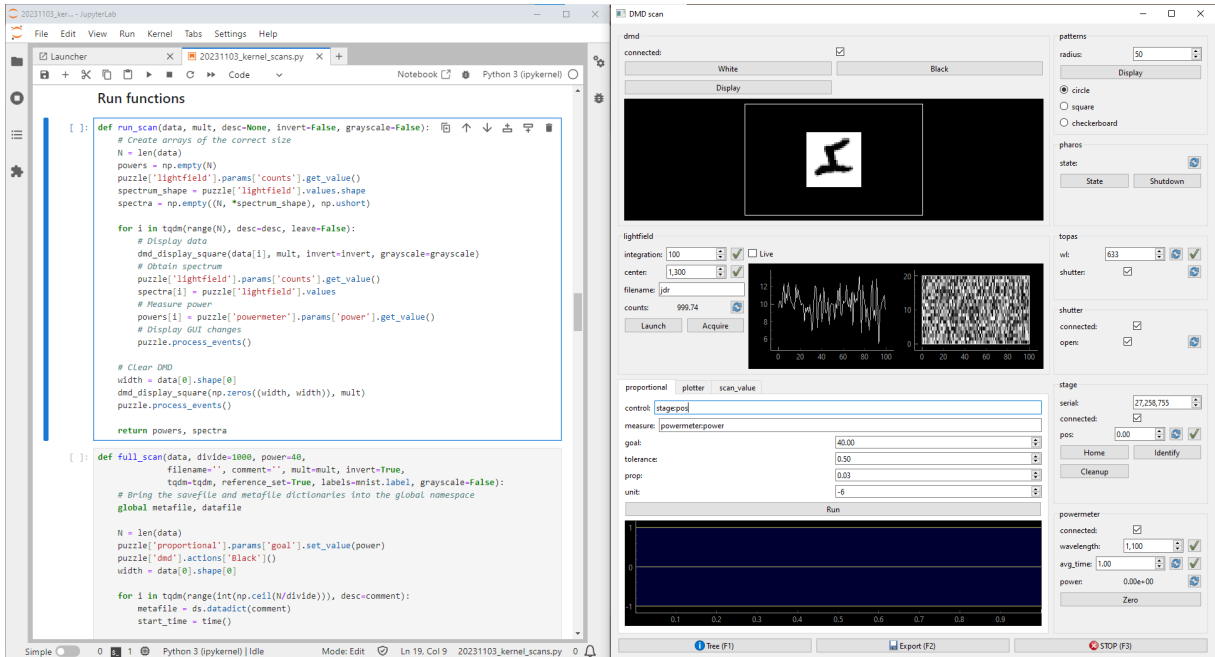
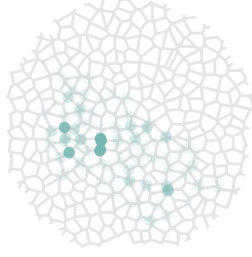


FIGURE A.1: A Jupyter Notebook running measurement code that directly manipulates an interactive puzzlepiece GUI.

Lastly, the design of puzzlepiece allows experiment automation to be highly modular. The code for interfacing with any given piece of hardware resides in a single module, referred to as a Piece, and multiple Pieces can be brought together to form a Puzzle – a GUI application. They can then communicate with each other through the Puzzle, all using the standardised

[API](#) introduced above. Different arrangements of Pieces can be easily created to suit different measurement needs, and everything can be operated from a single [GUI](#) window or a single Python script. Pieces that automate specific measurements can also be created, utilising the standard components provided by the library for rapid prototyping.

With the unified [API](#), automatic [GUI](#) generation, and high modularity, puzzlepiece has enabled many of the measurements presented in this thesis. I have been working with other students during my studies to improve the module itself, as well as its documentation and onboarding, with the hope that it can become an easy to use way to make coding for the lab better for everyone.



Appendix B

Additional figures

These figures provide additional details for topics discussed in the main text. Each caption provides a link back to the relevant section, and some details about the figure. Figures B.4, B.5, B.6, B.7, and B.8 are adapted from the following publication, and were produced by me, A. Fischer, and W. K. Ng:

J. Dranczewski,* W. K. Ng,* A. Fischer,* T. V. Raziman, D. Saxena, T. Farchy, K. Stenning, J. Peters, H. Schmid, W. R. Branford, M. Barahona, K. Moselund, R. Sapienza, J. C. Gartside, “[Retinomorphic Machine Vision in a Network Laser](#),” *arXiv*, Aug. 2024. [*manuscript in review*]

*Authors contributed equally.

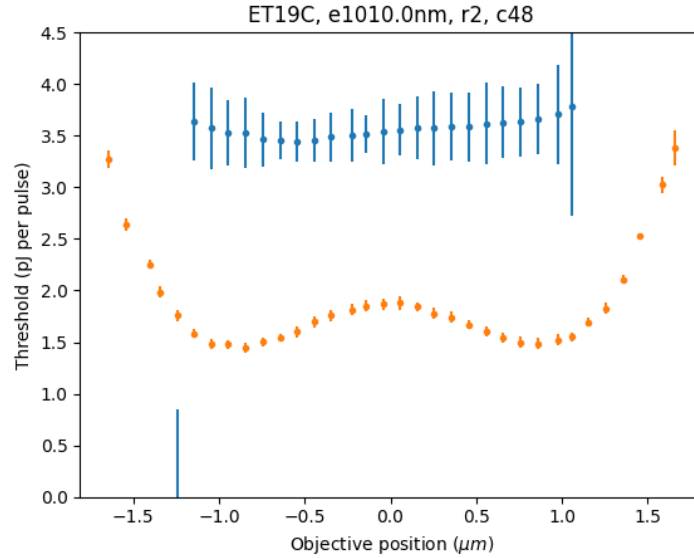


FIGURE B.1: Microdisk lasing threshold dependence on pump position (expanding on Section 3.3). The WGM in a disk exists around the edges, and so pumping close to the edge results in a higher pump efficiency and a lower threshold. Here, the lasing thresholds of two lasing peaks (marked orange and blue) in a microdisk device are tracked as the pump spot is swept across it, experimentally showing this effect. Automatically aligning to the centre of the disk is much more reliable, and so I have elected to do that despite the performance dip, which is shown here to be relatively small, and likely smaller than the experimental uncertainty introduced otherwise.

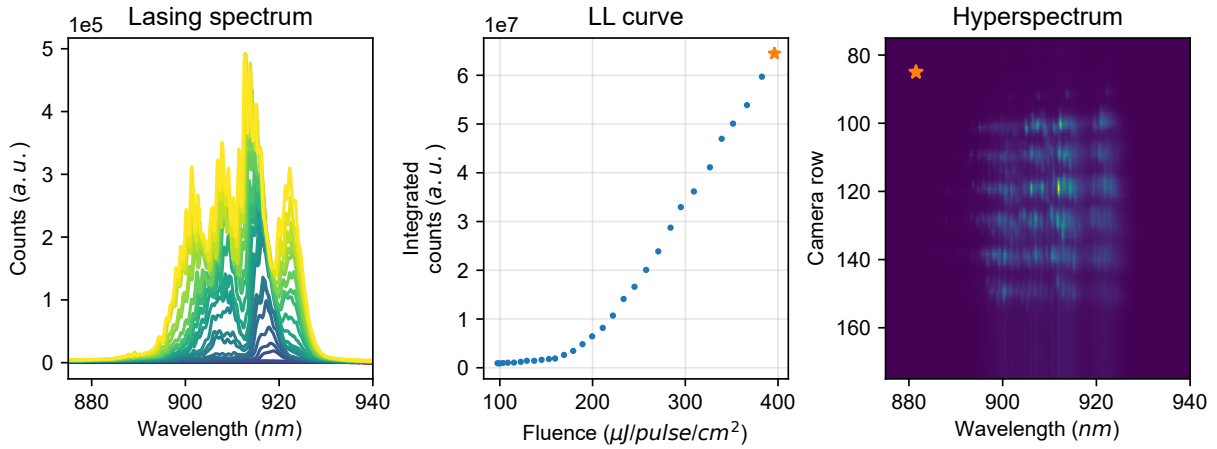


FIGURE B.2: 100 μm diameter, 10 μm average link length, 350 nm waveguide width network lasing with a ~ 70 ps pulse pump laser (expanding on Sections 4.1 and 6.2.1). The device does lase, but the threshold fluence is higher than for a femtosecond pulse. The lasing peaks are notably narrower, as expected from the chirp-based peak broadening mechanism explained in Section 3.3 – a picosecond pump pulse is closer to CW lasing, and the chirping thus less pronounced.

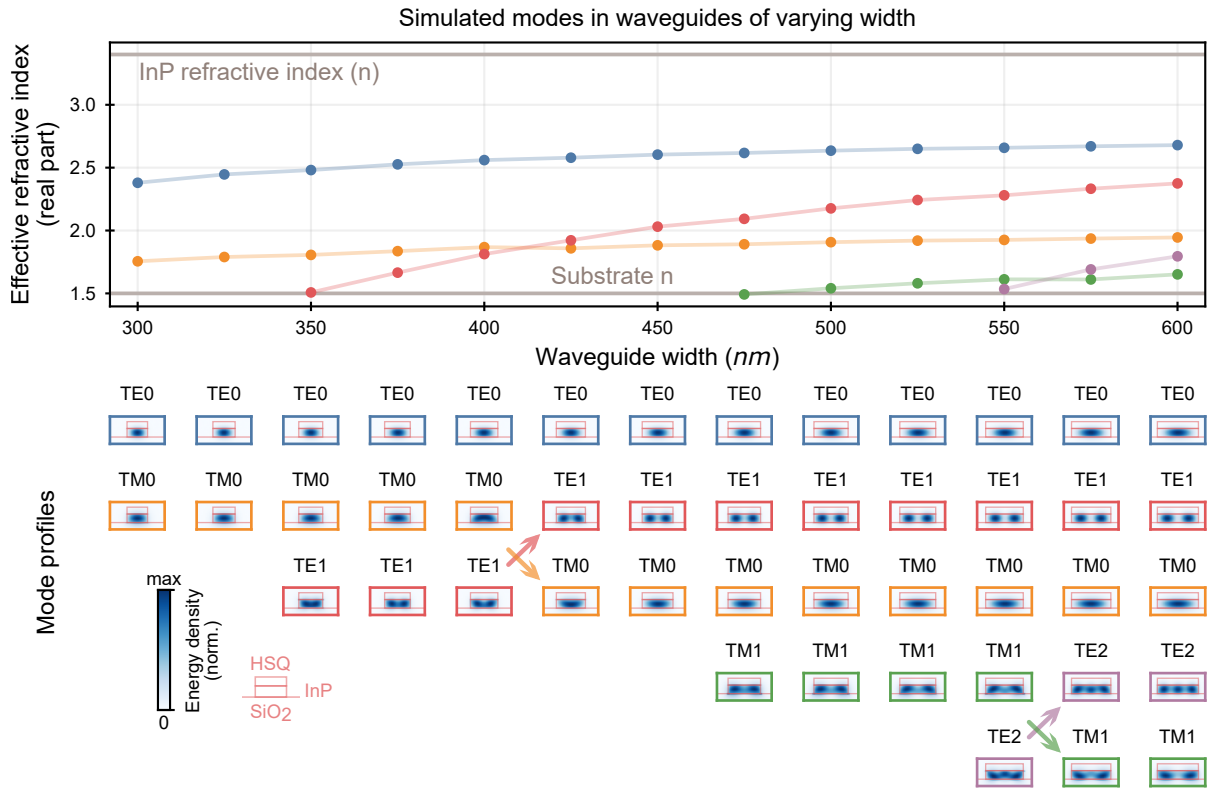


FIGURE B.3: Expanding on Fig. 4.5, mode profiles supported inside of waveguides with a more granular range of widths, with the matching effective refractive index shown above. Simulated with a FDE solver for light at 880 nm.



FIGURE B.4: Black and white version of the cameraman image shown in Fig. 5.1. Adapted from the original image (Ref. [168, 169], CC BY-NC 4.0) for our edge detection scheme, which requires black and white sections of the image to be displayed on the DMD.

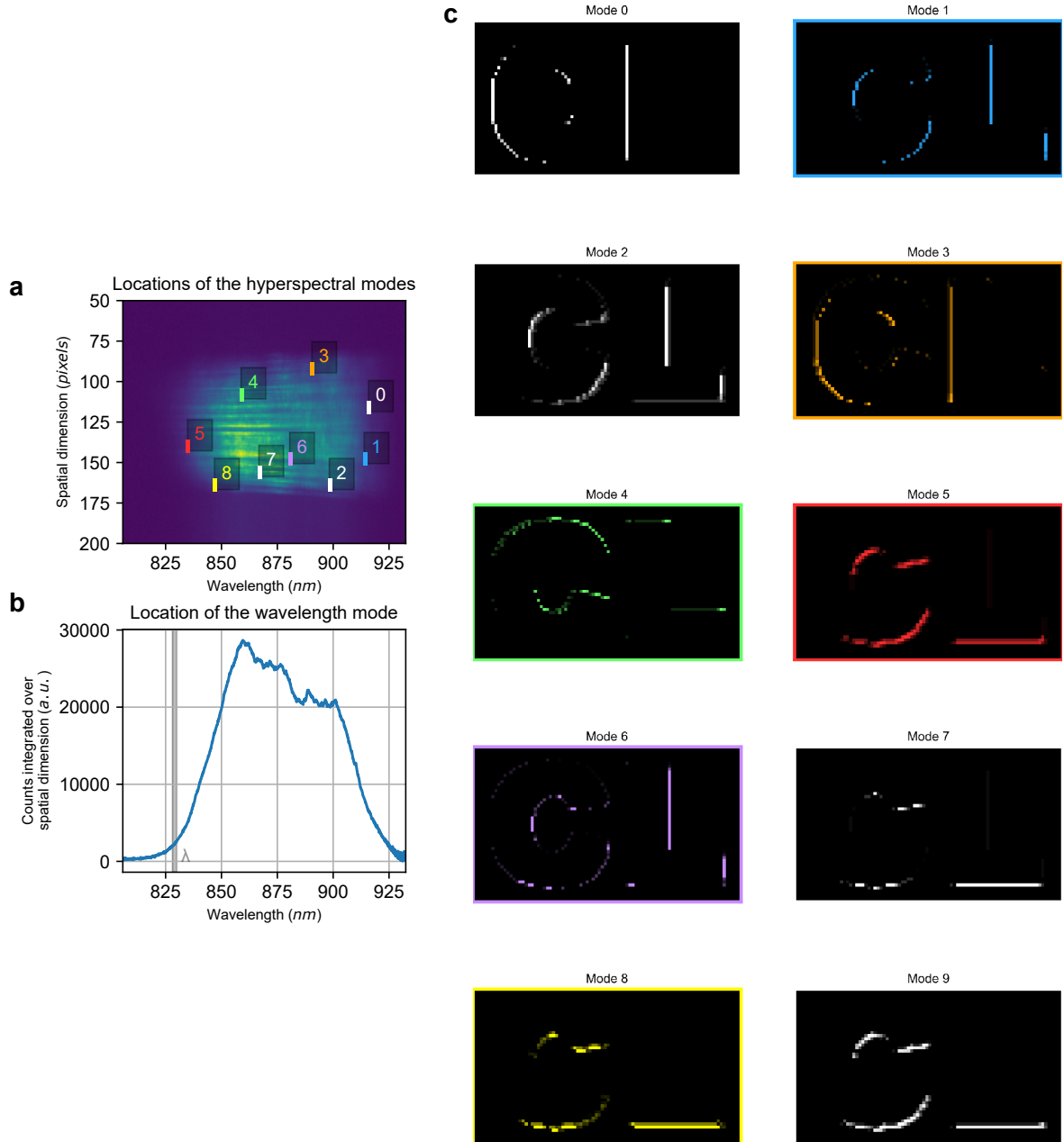


FIGURE B.5: Mode and feature maps for all library modes, including the spectral mode obtained by summing the hyperspectra along the spatial axis (expanding on Fig. 5.4). **a** The locations of the hyperspectral regions designated as feature-detecting modes. **b** The location of the spectral region designated as a feature-detecting mode. **c** Feature maps of the ‘CL’ image for all ten modes, mode 9 being the spectral mode.

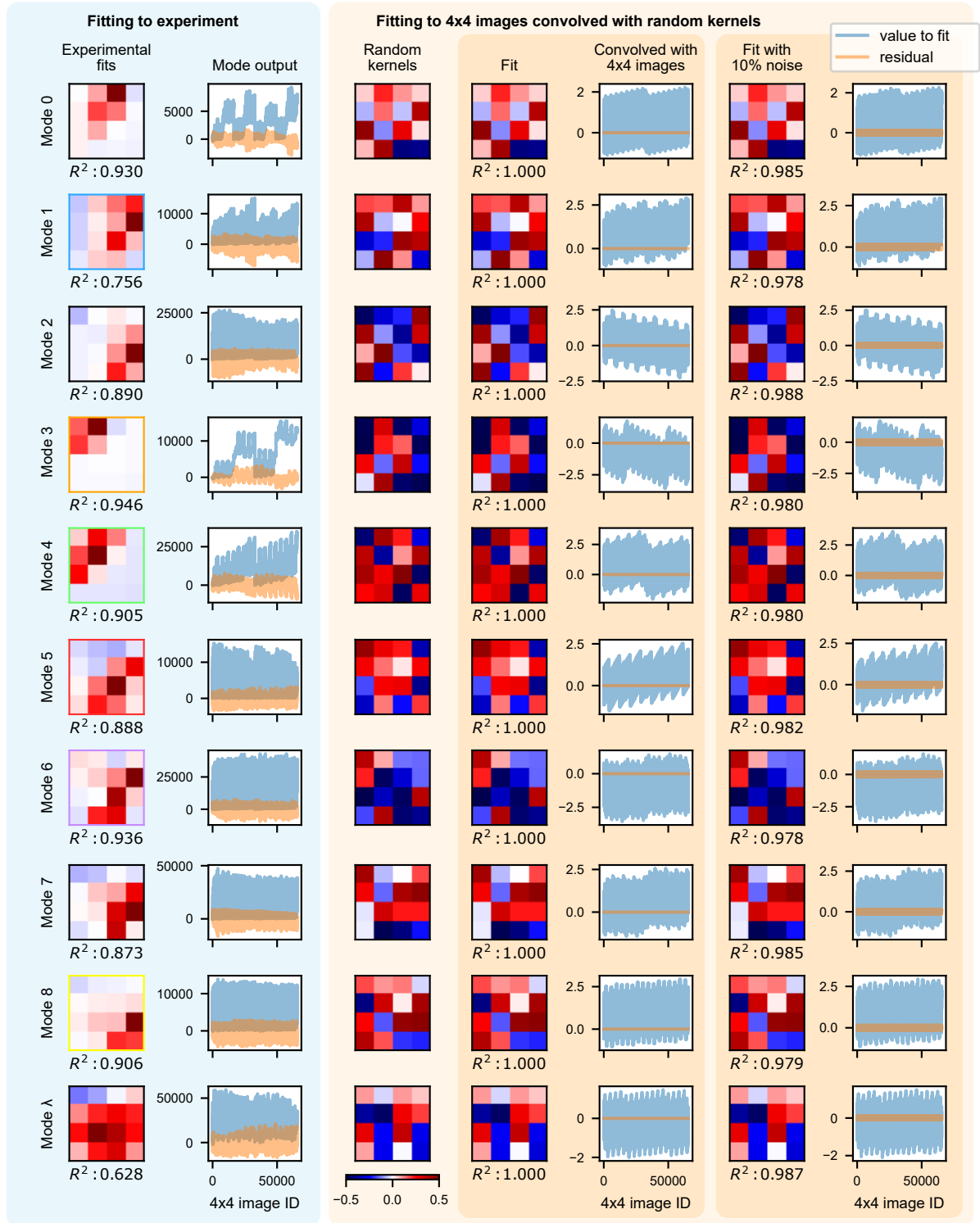


FIGURE B.6: As shown in Fig. 5.4, we can use least-squares fitting on the 4x4 image library hyperspectra to extract effective ‘kernels’ for the selected modes, which can be multiplied with the input image and summed to predict the mode output. This is not a perfect representation of the linear process, as indicated by the R^2 scores. To quantify whether the observed R^2 values could be a result of experimental noise rather than inherent non-linearity, I have performed a comparative study on randomly generated digital kernels. Convolution is a linear process, so the linear fit can reproduce it perfectly, and even adding random noise to the data does not bring R^2 values to the experimental levels.



FIGURE B.7: Examples of the ten classes of the MNIST dataset and the edges detected for each of them by the modes in our experimental 4x4 image hyperspectra library (expanding on Section 5.3). Modes presented in the main text are highlighted in colour. Mode 9 is the spectral mode, as defined in Fig. B.5. These edge maps would be projected onto the network laser in the multilayer scheme as shown in Fig. 5.8.



FIGURE B.8: Examples of the ten classes of the Fashion MNIST dataset and the edges detected for each of them by the modes in our experimental 4x4 image hyperspectra library (expanding on Section 5.4 and Rig. 5.9). Modes presented in the main text are highlighted in colour. Mode 9 is the spectral mode, as defined in Fig. B.5. These edge maps would be projected onto the network laser in the multilayer scheme as shown in Fig. 5.8, with Fashion MNIST results specifically in Fig. 5.9.



FIGURE B.9: Logistic regression weights for a two layer classifier, showing how different hyperspectral regions are weighted when predicting whether the spectrum belongs to a benign or malignant sample in BreakHis 400X (expanding on Section 5.4). Fully white regions were rejected during feature selection. The classifier utilises spectral features from all the different channel and mode map illuminations, showing that each one contributes to the final prediction.

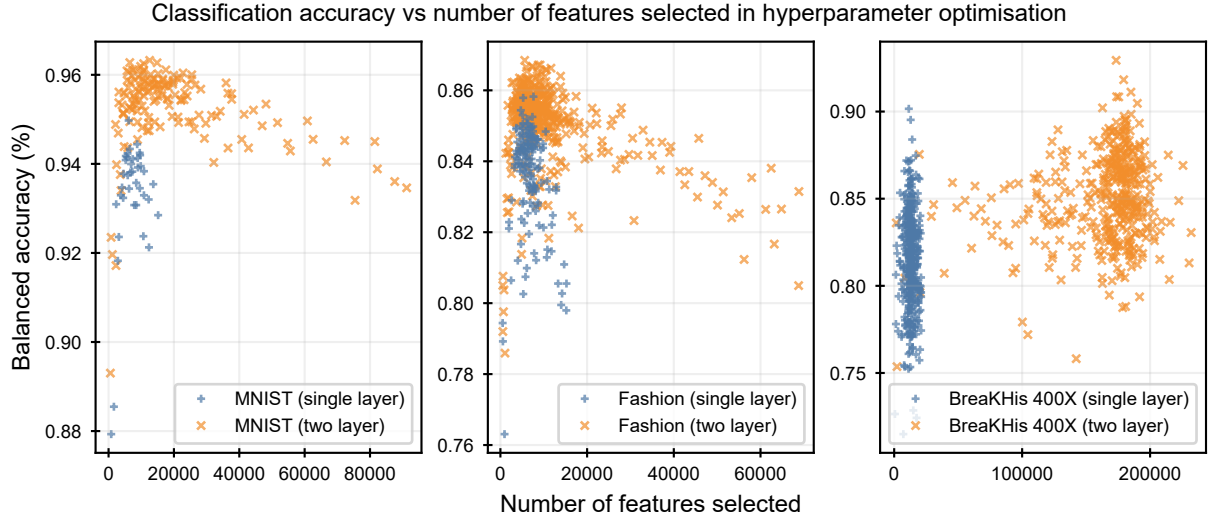
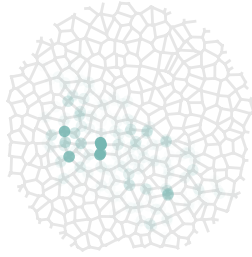


FIGURE B.10: As part of the hyperparameter optimisation process mentioned in Section 5.2, we vary the number of selected features and the amount of regularisation applied in the logistic regression fitting. This figure shows the classification accuracies obtained in the hyperparameter optimisation for our three main tasks as a function of the number of features used. We can compare this to the number of trainable parameters for the digital benchmarks shown in Fig. 5.11 (65k - 267k) – note that for logistic regression the number of weights is the number of features multiplied by the number of classes (10 for MNIST and Fashion MNIST, one for BreakHis). This means that the optimal number of weights falls in the region of 100k for the extreme learning machine approach, with more weights for the two layer BreakHis, which used 11 hyperspectra per RGB channel. The comparison comes with caveats though – looking at Appendix Fig. B.9, we can see that a large number of the features is not utilised, and the positive and negative weights occur in clusters that could be reduced to a single weight. Our feature selection algorithm is quite naive (to simplify the process and highlight that the physics does most of the computation), and so it is not very good at reducing the weight number – we believe that it should be possible to optimise much further than the 100k estimated here, and even with the simple algorithm we retain strong performance when fewer features are selected. The number of features is also not one-to-one comparable to the number of trainable parameters in neural networks, as logistic regression is generally less computationally expensive to train, not requiring backpropagation or non-linear activation.



Appendix C

Fabrication images

Nanofabrication can be finicky, but the results of mishaps at the nanoscale are often beautiful. This Appendix contains a small gallery of interesting images I have taken during cleanroom experiments.

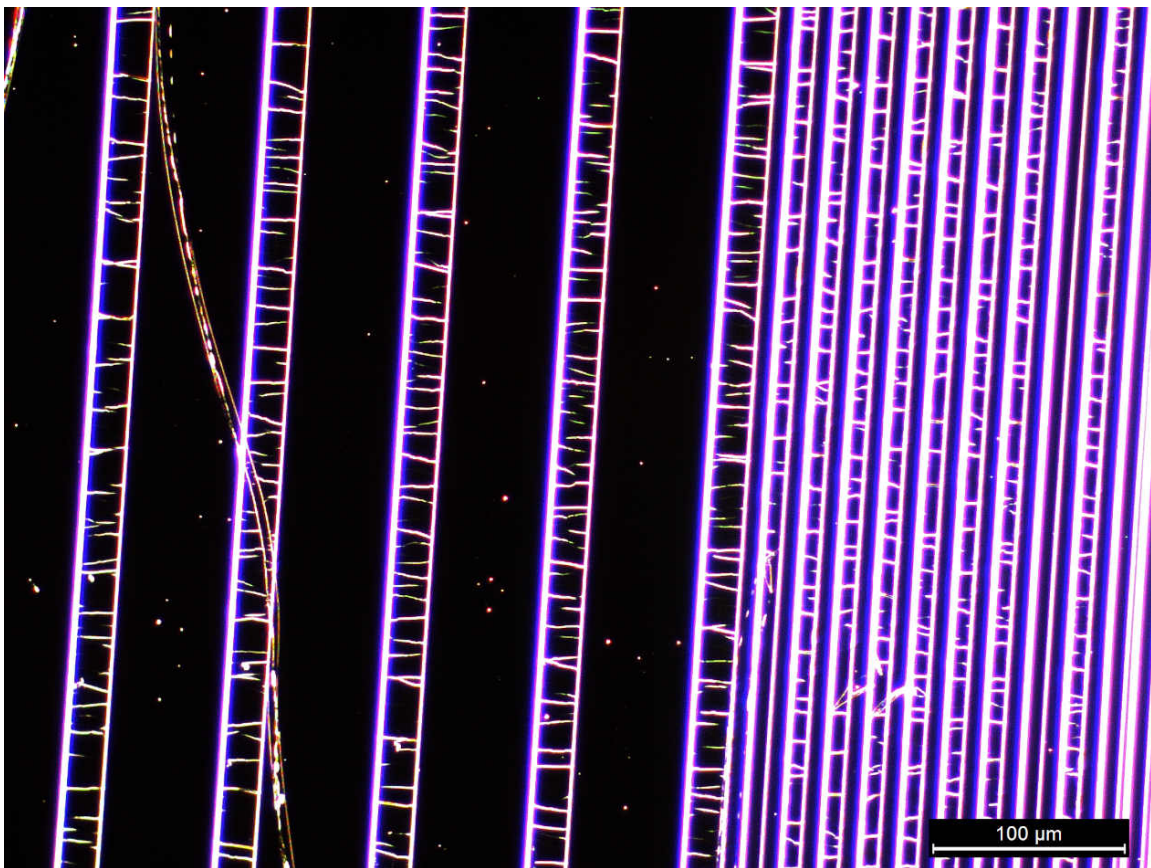


FIGURE C.1: Lines of cracked optical resist under dark-field microscopy. Working with optical resist involves multiple heat treatment step, and this chip did not handle them well. The cracks are enhanced by the dark-field image, which collects only the light scattered from structures and surface imperfections.

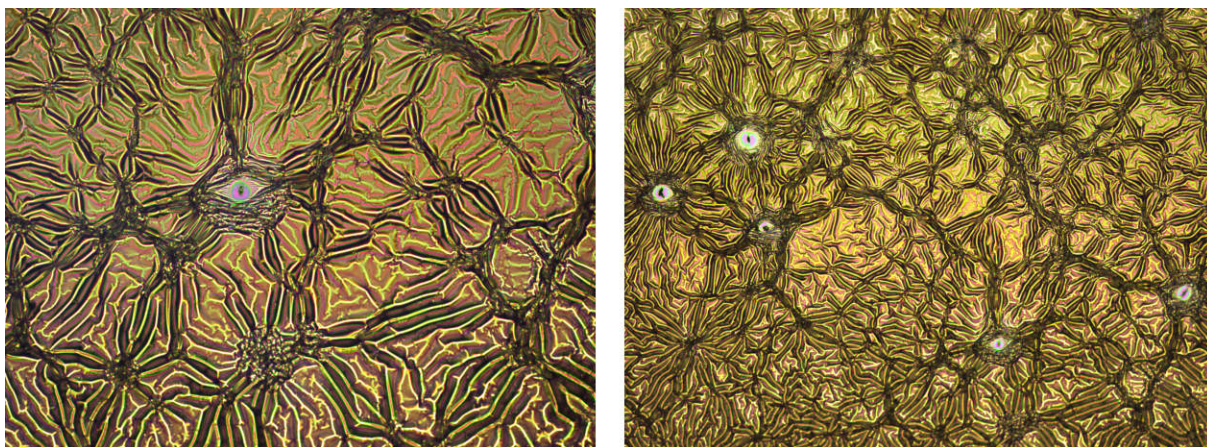


FIGURE C.2: *Plasma-enhanced chemical vapour deposition (PECVD) SiO_2 deposited on top of a hardened optical mask layer, resulting in bad bonding and an uneven surface. The optical mask was protecting the chip in storage and should have been removed prior to the deposition.*

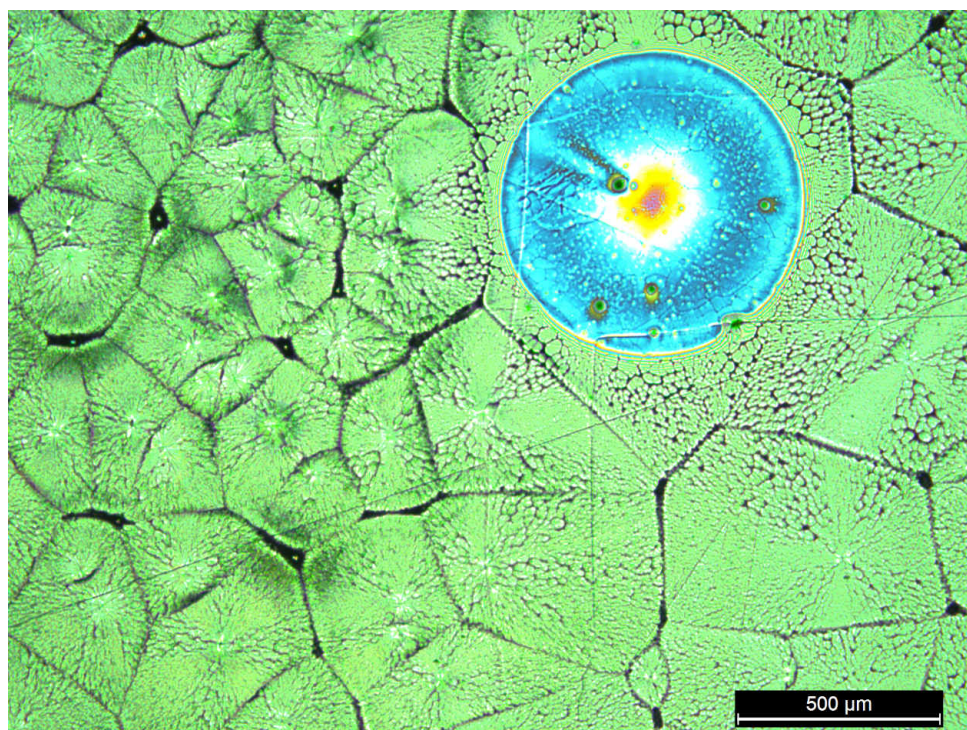


FIGURE C.3: Wax is used to bond chips to ICP carrier wafers, and it sometimes developed bubbles when exposed to the plasma.

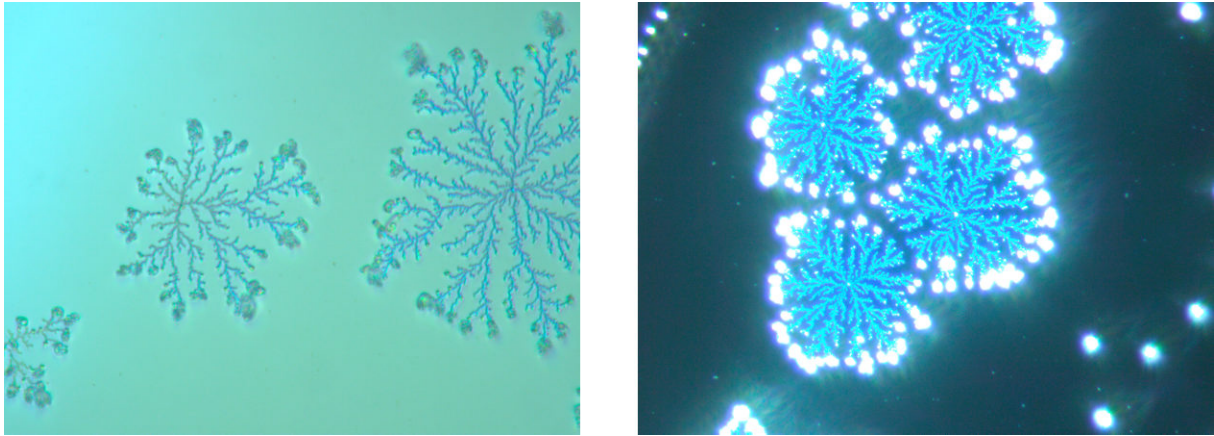


FIGURE C.4: A plasma is a challenging environment, and on rare occasions the silicon carrier wafer gets slightly damaged. In this cases discharges from the plasma left tree-like Lichtenberg figures, shown here in bright-field (left) and dark-field (right) microscopy.

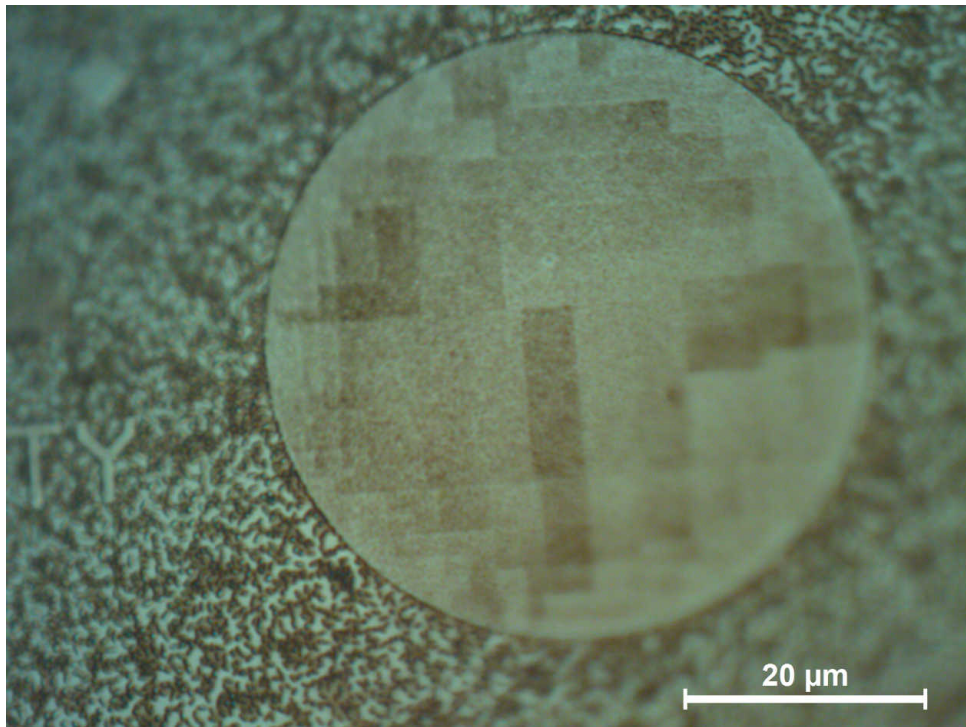


FIGURE C.5: This failed e-beam resist development showcases the way electron beam lithography writes large curved structures – by dividing them into rectangles of various sizes, which can be more easily scanned with the beam. The written rectangles are visible on the large disk here.

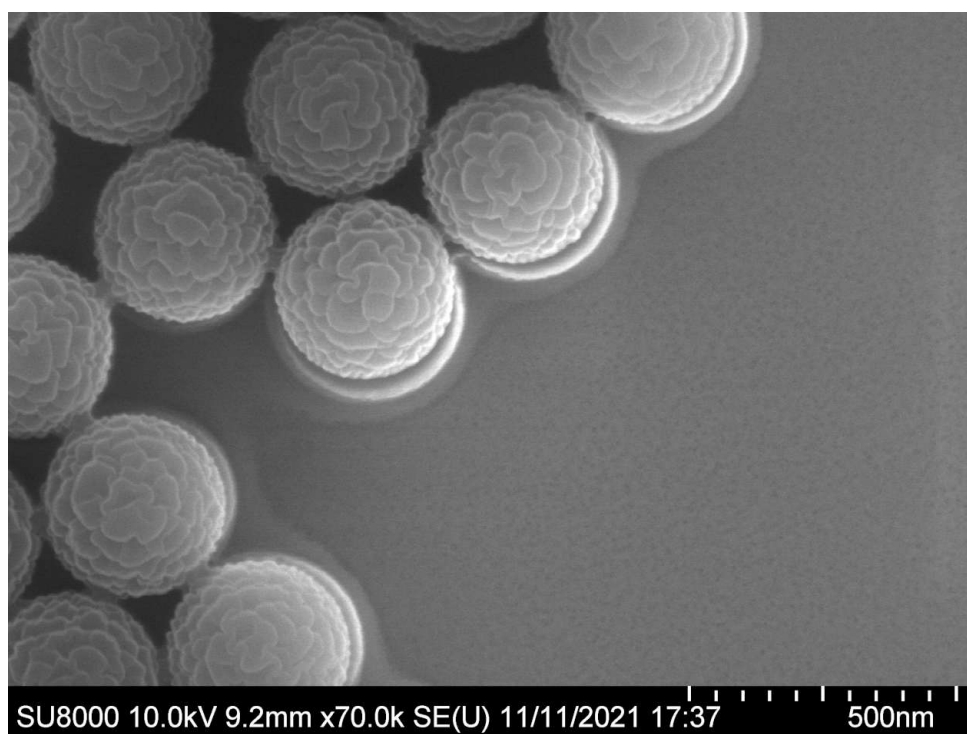


FIGURE C.6: I have used polymer nano-spheres as a soft mask for the very first etch tests. The polymer does not stand up well to the plasma environment, degrading into the floral patterns shown in this SEM image.



FIGURE C.7: The soft mask nano-spheres were deposited from a colloidal solution, meaning that the deposition patterns occasionally followed the flow of the liquid. Here, a dark-field microscope image shows the spheres clustering around a piece of debris – a starry night.

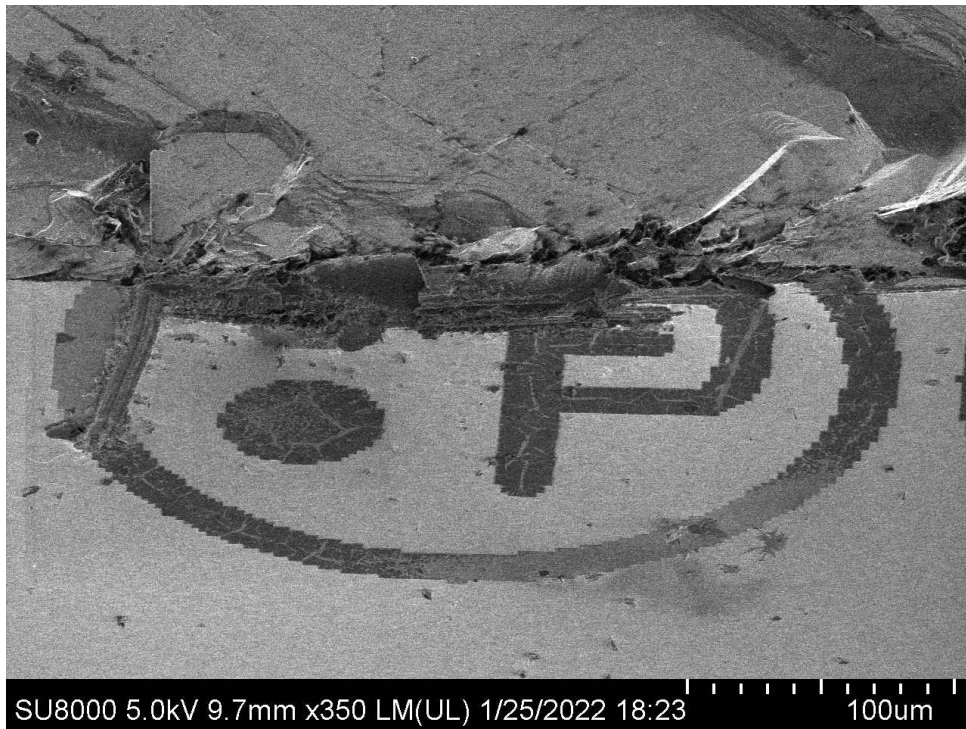


FIGURE C.8: Cleaving normally produces smooth edges along the crystal planes of the chip, but in this SEM image it did not go so cleanly. The cleave goes through an optical mask image – a test structure showing an emoticon.

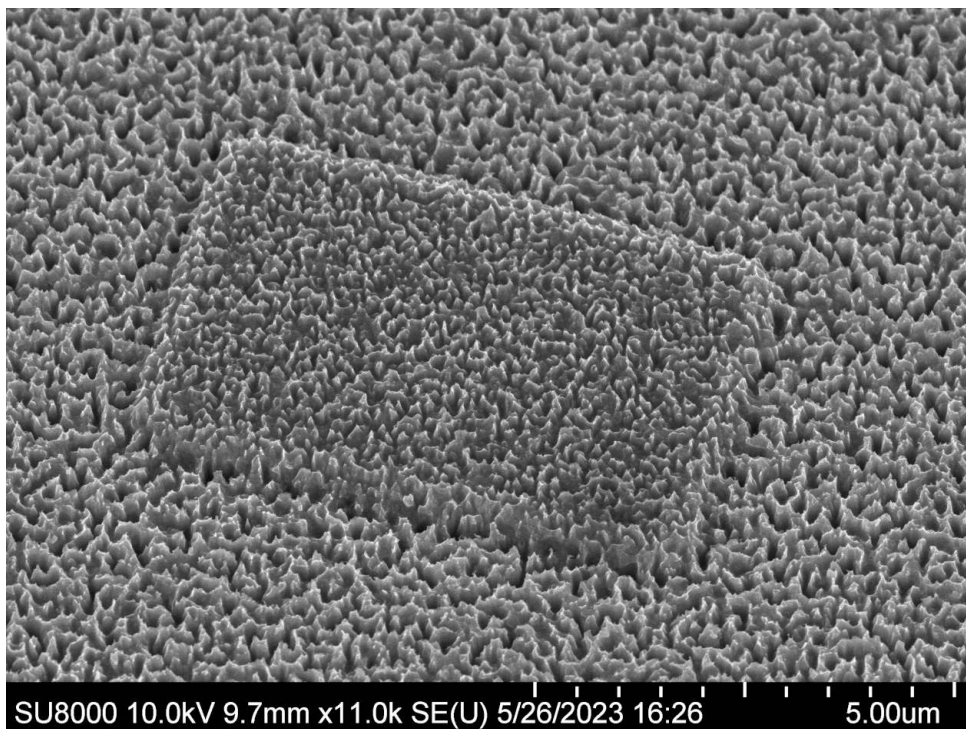


FIGURE C.9: The etch test in this SEM image failed in multiple ways. First, the surface of the material is highly uneven. Second, the selectivity for this recipe was so low, that it completely etched through the mask, exposing the bump of material in the centre of the image.

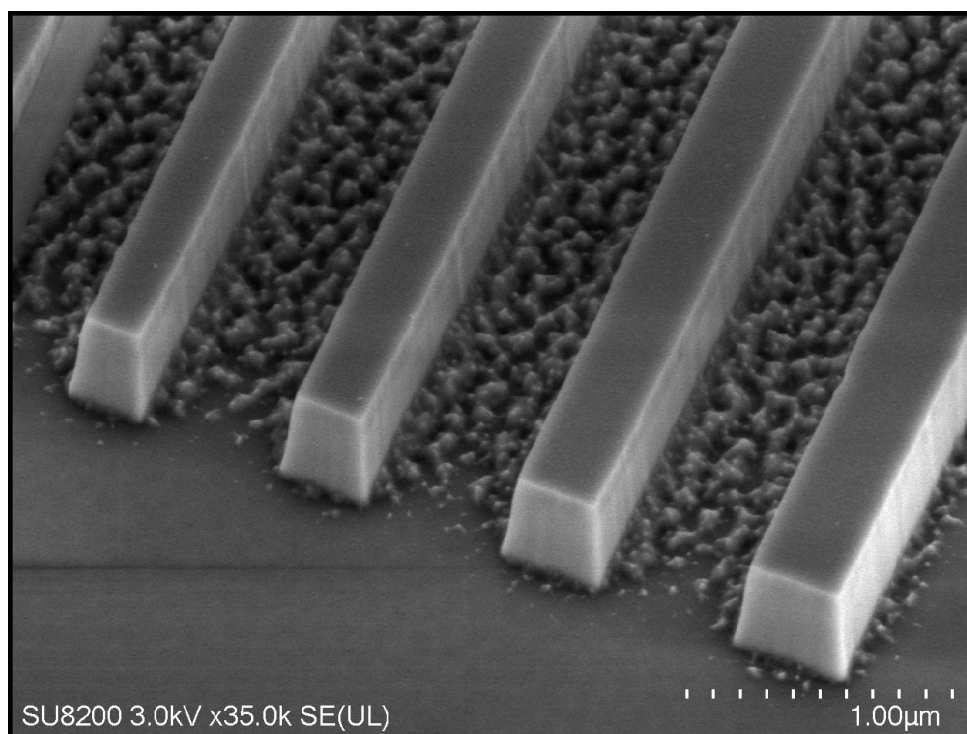


FIGURE C.10: Plasma etch ‘loading effects’ mean that etching is faster for unobstructed areas, and slower next to devices. This continuous etch was not run long enough, and so some material remains in the spaces between waveguides, despite the etch going completely through the bonded InP layer in the areas further from the mask.

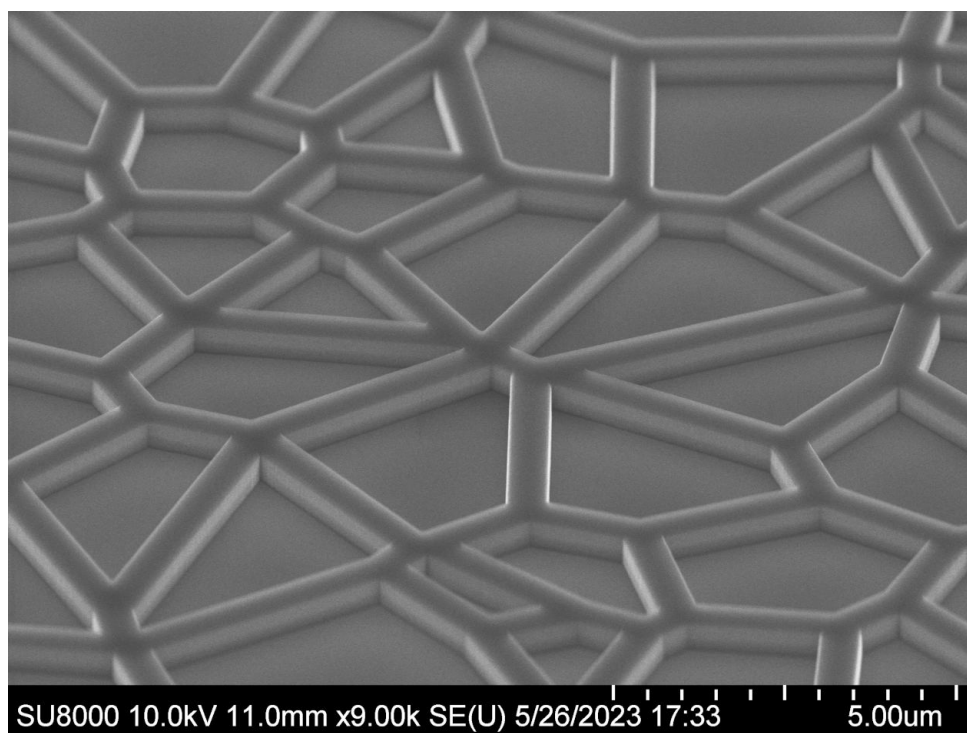


FIGURE C.11: When things go right, fabrication can be beautiful too. This SEM image shows a side-view of a cyclic-etched InP network laser with embedded horizontal QWs.

Reproduction permissions

When necessary, the appropriate license and a link to its full text have been provided directly next to any reproduced figures, or in the originality statement in the chapter heading. For print copies of the thesis, the links to the relevant licenses are reproduced verbatim here:

- CC BY 4.0 – <https://creativecommons.org/licenses/by/4.0/>
- CC BY-NC 4.0 – <https://creativecommons.org/licenses/by-nc/4.0/>

Fig. 1.1 **b** is adapted from Ref. [8] (© 2017 IEEE), under the following condition: *In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.*

For Fig. 1.3, the data and figures were made available under the MIT License, which is reproduced below:

MIT License

Copyright (c) 2021 Amir Gholami

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Bibliography

- [1] G. E. Moore, “Cramming more components onto integrated circuits,” *Electronics*, vol. 38, no. 8, 1965.
- [2] G. E. Moore, “Progress in digital integrated electronics [Technical literature, Copyright 1975 IEEE. Reprinted, with permission. Technical Digest. International Electron Devices Meeting, IEEE, 1975, pp. 11-13.],” *IEEE Solid-State Circuits Society Newsletter*, vol. 11, no. 3, pp. 36–37, Sep. 2006.
- [3] I. Tuomi, “The Lives and Death of Moore’s Law,” *First Monday*, Nov. 2002.
- [4] P. A. Packan, “Pushing the Limits,” *Science*, vol. 285, no. 5436, pp. 2079–2081, Sep. 1999.
- [5] E. Pop, S. Sinha, and K. Goodson, “Heat Generation and Transport in Nanometer-Scale Transistors,” *Proceedings of the IEEE*, vol. 94, no. 8, pp. 1587–1601, Aug. 2006.
- [6] K. S. Kim, J. Kwon, H. Ryu *et al.*, “The future of two-dimensional semiconductors beyond Moore’s law,” *Nature Nanotechnology*, vol. 19, no. 7, pp. 895–906, Jul. 2024.
- [7] “Karlrupp/microprocessor-trend-data: Data repository for my blog series on microprocessor trend data.” Feb. 2022. [Online]. Available: <https://github.com/karlrupp/microprocessor-trend-data> (Accessed: 2025-02-20)
- [8] T. N. Theis and H.-S. P. Wong, “The End of Moore’s Law: A New Beginning for Information Technology,” *Computing in Science & Engineering*, vol. 19, no. 2, pp. 41–50, Mar. 2017.
- [9] A. Gholami, Z. Yao, S. Kim *et al.*, “AI and Memory Wall,” *IEEE Micro*, vol. 44, no. 03, pp. 33–39, May 2024.
- [10] A. Radford, J. Wu, R. Child *et al.*, “Language Models are Unsupervised Multitask Learners,” *OpenAI*, 2019.
- [11] S. Chen, “Data centres will use twice as much energy by 2030 — driven by AI,” *Nature*, Apr. 2025.
- [12] S. Chen, “How much energy will AI really consume? The good, the bad and the unknown,” *Nature*, vol. 639, no. 8053, pp. 22–24, Mar. 2025.
- [13] A. S. Tanenbaum, *Structured Computer Organization*, 6th ed. Harlow, England: Pearson, 2013.
- [14] J. von Neumann, “First draft of a report on the EDVAC,” *IEEE Annals of the History of Computing*, vol. 15, no. 4, pp. 27–75, 1993.
- [15] “Big data needs a hardware revolution.” [Online]. Available: <https://www.nature.com/articles/d41586-018-01683-1> (Accessed: 2025-05-20)

-
- [16] D. A. B. Miller, “Attojoule Optoelectronics for Low-Energy Information Processing and Communications,” *Journal of Lightwave Technology*, vol. 35, no. 3, pp. 346–396, Feb. 2017.
 - [17] Z. Li, F. Liu, W. Yang *et al.*, “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022.
 - [18] J. Gu, Z. Wang, J. Kuen *et al.*, “Recent advances in convolutional neural networks,” *Pattern Recognition*, vol. 77, pp. 354–377, May 2018.
 - [19] Y. LeCun, L. D. Jackel, L. Bottou *et al.*, “Learning algorithms for classification: A comparison on handwritten digit recognition,” *Neural networks: the statistical mechanics perspective*, vol. 261, no. 276, p. 2, 1995.
 - [20] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” *arXiv*, Dec. 2015. [Online]. Available: <http://arxiv.org/abs/1511.08458> (Accessed: 2025-04-29)
 - [21] K. Chellapilla, S. Puri, and P. Simard, “High Performance Convolutional Neural Networks for Document Processing,” in *Tenth International Workshop on Frontiers in Handwriting Recognition*. Suvisoft, 2006. [Online]. Available: <https://inria.hal.science/inria-00112631v1/file/p1038112283956.pdf> (Accessed: 2025-04-29)
 - [22] T.-J. Yang, Y.-H. Chen, J. Emer *et al.*, “A method to estimate the energy consumption of deep neural networks,” in *2017 51st Asilomar Conference on Signals, Systems, and Computers*, Oct. 2017, pp. 1916–1920. [Online]. Available: <https://ieeexplore.ieee.org/document/8335698> (Accessed: 2025-05-16)
 - [23] A. Gholami, “Amirgholami/ai_and_memory_wall,” Apr. 2025. [Online]. Available: https://github.com/amirgholami/ai_and_memory_wall (Accessed: 2025-04-07)
 - [24] Y. Song, B. Millidge, T. Salvatori *et al.*, “Inferring neural activity before plasticity as a foundation for learning beyond backpropagation,” *Nature Neuroscience*, vol. 27, no. 2, pp. 348–358, Feb. 2024.
 - [25] V. Balasubramanian, “Brain power,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 32, p. e2107022118, Aug. 2021.
 - [26] A. Momeni, B. Rahmani, B. Scellier *et al.*, “Training of Physical Neural Networks,” *arXiv*, Jun. 2024. [Online]. Available: <http://arxiv.org/abs/2406.03372> (Accessed: 2025-05-18)
 - [27] A. Litvinenko, R. Khymyn, V. H. González *et al.*, “A spinwave Ising machine,” *Communications Physics*, vol. 6, no. 1, p. 227, Aug. 2023.
 - [28] D. Marković, A. Mizrahi, D. Querlioz *et al.*, “Physics for neuromorphic computing,” *Nature Reviews Physics*, vol. 2, no. 9, pp. 499–510, Jul. 2020.
 - [29] S. Kumar, J. P. Strachan, and R. S. Williams, “Chaotic dynamics in nanoscale NbO₂ Mott memristors for analogue computing,” *Nature*, vol. 548, no. 7667, pp. 318–321, Aug. 2017.
 - [30] J. Robertson, M. Hejda, J. Bueno *et al.*, “Ultrafast optical integration and pattern classification for neuromorphic photonics based on spiking VCSEL neurons,” *Scientific Reports*, vol. 10, no. 1, p. 6098, Apr. 2020.
 - [31] T. Ohno, T. Hasegawa, T. Tsuruoka *et al.*, “Short-term plasticity and long-term potentiation mimicked in single inorganic synapses,” *Nature Materials*, vol. 10, no. 8, pp. 591–595, Aug. 2011.

- [32] M. Ignatov, M. Ziegler, M. Hansen *et al.*, “Memristive stochastic plasticity enables mimicking of neural synchrony: Memristive circuit emulates an optical illusion,” *Science Advances*, vol. 3, no. 10, p. e1700849, Oct. 2017.
- [33] P. Yao, H. Wu, B. Gao *et al.*, “Fully hardware-implemented memristor convolutional neural network,” *Nature*, vol. 577, no. 7792, pp. 641–646, Jan. 2020.
- [34] A. Singh, S. Diware, A. Gebregiorgis *et al.*, “Low-Power Memristor-Based Computing for Edge-AI Applications,” in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2021, pp. 1–5. [Online]. Available: <https://ieeexplore.ieee.org/document/9401226/authors> (Accessed: 2025-05-17)
- [35] P. M. Solomon, “Analog neuromorphic computing using programmable resistor arrays,” *Solid-State Electronics*, vol. 155, pp. 82–92, May 2019.
- [36] T. Wang, S.-Y. Ma, L. G. Wright *et al.*, “An optical neural network using less than 1 photon per multiplication,” *Nature Communications*, vol. 13, no. 1, p. 123, Jan. 2022.
- [37] M. G. Anderson, S.-Y. Ma, T. Wang *et al.*, “Optical Transformers,” *arXiv*, Feb. 2023. [Online]. Available: <http://arxiv.org/abs/2302.10360> (Accessed: 2025-05-17)
- [38] H. Zhou, J. Dong, J. Cheng *et al.*, “Photonic matrix multiplication lights up photonic accelerator and beyond,” *Light: Science & Applications*, vol. 11, no. 1, p. 30, Feb. 2022.
- [39] X. Lin, Y. Rivenson, N. T. Yardimci *et al.*, “All-optical machine learning using diffractive deep neural networks,” *Science*, vol. 361, no. 6406, pp. 1004–1008, Sep. 2018.
- [40] M. Yildirim, N. U. Dinc, I. Oguz *et al.*, “Nonlinear processing with linear optics,” *Nature Photonics*, vol. 18, no. 10, pp. 1076–1082, Oct. 2024.
- [41] H. Zheng, Q. Liu, I. I. Kravchenko *et al.*, “Multichannel meta-imagers for accelerating machine vision,” *Nature Nanotechnology*, vol. 19, no. 4, pp. 471–478, Apr. 2024.
- [42] Y. Shen, N. C. Harris, S. Skirlo *et al.*, “Deep learning with coherent nanophotonic circuits,” *Nature Photonics*, vol. 11, no. 7, pp. 441–446, Jul. 2017.
- [43] A. Sludds, S. Bandyopadhyay, Z. Chen *et al.*, “Delocalized photonic deep learning on the internet’s edge,” *Science*, vol. 378, no. 6617, pp. 270–276, Oct. 2022.
- [44] J. Feldmann, N. Youngblood, M. Karpov *et al.*, “Parallel convolutional processing using an integrated photonic tensor core,” *Nature*, vol. 589, no. 7840, pp. 52–58, Jan. 2021.
- [45] I. Oguz, J.-L. Hsieh, N. U. Dinc *et al.*, “Programming nonlinear propagation for efficient optical learning machines,” *Advanced Photonics*, vol. 6, no. 1, p. 016002, Jan. 2024.
- [46] J. C. Gartside, K. D. Stenning, A. Vanstone *et al.*, “Reconfigurable training and reservoir computing in an artificial spin-vortex ice via spin-wave fingerprinting,” *Nature Nanotechnology*, vol. 17, no. 5, pp. 460–469, May 2022.
- [47] J. Torrejon, M. Riou, F. A. Araujo *et al.*, “Neuromorphic computing with nanoscale spintronic oscillators,” *Nature*, vol. 547, no. 7664, pp. 428–431, Jul. 2017.
- [48] M. Yan, C. Huang, P. Bienstman *et al.*, “Emerging opportunities and challenges for the future of reservoir computing,” *Nature Communications*, vol. 15, no. 1, p. 2056, Mar. 2024.

-
- [49] D. Brunner, M. C. Soriano, C. R. Mirasso *et al.*, “Parallel photonic information processing at gigabyte per second data rates using transient states,” *Nature Communications*, vol. 4, no. 1, p. 1364, Jan. 2013.
 - [50] U. Teğın, M. Yıldırım, İ. Oğuz *et al.*, “Scalable optical learning operator,” *Nature Computational Science*, vol. 1, no. 8, pp. 542–549, Aug. 2021.
 - [51] L. G. Wright, T. Onodera, M. M. Stein *et al.*, “Deep physical neural networks trained with backpropagation,” *Nature*, vol. 601, no. 7894, pp. 549–555, Jan. 2022.
 - [52] X. Porte, A. Skalli, N. Haghighi *et al.*, “A complete, parallel and autonomous photonic neural network in a semiconductor multimode laser,” *Journal of Physics: Photonics*, vol. 3, no. 2, p. 024017, Apr. 2021.
 - [53] J. C. Coulombe, M. C. A. York, and J. Sylvestre, “Computing with networks of nonlinear mechanical oscillators,” *PLOS ONE*, vol. 12, no. 6, p. e0178663, Jun. 2017.
 - [54] D. S. Wiersma, “The physics and applications of random lasers,” *Nature Physics*, vol. 4, no. 5, pp. 359–367, May 2008.
 - [55] R. Sapienza, “Determining random lasing action,” *Nature Reviews Physics*, vol. 1, no. 11, pp. 690–695, Nov. 2019.
 - [56] J. G. Câmara, D. M. da Silva, L. R. P. Kassab *et al.*, “Random laser emission from neodymium doped zinc tellurite glass-powder presenting luminescence concentration quenching,” *Journal of Luminescence*, vol. 233, p. 117936, May 2021.
 - [57] P. D. García, R. Sapienza, Á. Blanco *et al.*, “Photonic Glass: A Novel Random Material for Light,” *Advanced Materials*, vol. 19, no. 18, pp. 2597–2602, 2007.
 - [58] S. Gottardo, S. Cavalieri, O. Yaroshchuk *et al.*, “Quasi-Two-Dimensional Diffusive Random Laser Action,” *Physical Review Letters*, vol. 93, no. 26, p. 263901, Dec. 2004.
 - [59] F. J. P., Schuurmans, D. Vanmaekelbergh *et al.*, “Strongly Photonic Macroporous Gallium Phosphide Networks,” *Science*, vol. 284, no. 5411, pp. 141–143, Apr. 1999.
 - [60] B. Kumar, R. Homri, and P. Sebbah, “2D tunable all-solid-state random laser in the visible,” *Scientific Reports*, vol. 13, no. 1, May 2023.
 - [61] H. Noh, J.-K. Yang, S. F. Liew *et al.*, “Photonic network laser,” *Optics Letters*, vol. 36, no. 18, pp. 3560–3562, Sep. 2011.
 - [62] M. Trivedi, D. Saxena, W. K. Ng *et al.*, “Self-organized lasers from reconfigurable colloidal assemblies,” *Nature Physics*, vol. 18, no. 8, pp. 939–944, Aug. 2022.
 - [63] M. Gaio, D. Saxena, J. Bertolotti *et al.*, “A nanophotonic laser on a graph,” *Nature Communications*, vol. 10, no. 1, p. 226, Jan. 2019.
 - [64] J. Xue, T. Wu, Y. Dai *et al.*, “Electrospinning and Electrospun Nanofibers: Methods, Materials, and Applications,” *Chemical reviews*, vol. 119, no. 8, pp. 5298–5415, Apr. 2019.
 - [65] D. Saxena, A. Arnaudon, O. Cipolato *et al.*, “Sensitivity and spectral control of network lasers,” *Nature Communications*, vol. 13, no. 1, p. 6493, Oct. 2022.

BIBLIOGRAPHY

- [66] H. E. Türeci, A. D. Stone, and B. Collier, “Self-consistent multimode lasing theory for complex or random lasing media,” *Physical Review A*, vol. 74, no. 4, p. 043822, Oct. 2006.
- [67] W. K. Ng, J. Dranczewski, A. Fischer *et al.*, “Few-Shot Retinomorphic Vision in a Nonlinear Photonic Network Laser,” *arXiv*, Jun. 2025. [Online]. Available: <http://arxiv.org/abs/2407.15558> (Accessed: 2025-09-22)
- [68] A. Bett, F. Dimroth, G. Stollwerck *et al.*, “III-V compounds for solar cell applications,” *Applied Physics A*, vol. 69, no. 2, pp. 119–129, Aug. 1999.
- [69] G. Crosnier, D. Sanchez, S. Bouchoule *et al.*, “Hybrid indium phosphide-on-silicon nanolaser diode,” *Nature Photonics*, vol. 11, no. 5, pp. 297–300, May 2017.
- [70] G. Morthier, T. Spuesens, P. Mechet *et al.*, “InP Microdisk Lasers Integrated on Si for Optical Interconnects,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 21, no. 6, pp. 359–368, Nov. 2015.
- [71] T. Kakitsuka and S. Matsuo, “Current-injection Photonic-crystal Laser,” *NTT Technical Review*, vol. 10, no. 12, p. 6, 2012.
- [72] T. Ferrotti, B. Blampey, C. Jany *et al.*, “Co-integrated 1.3 μ m hybrid III-V/silicon tunable laser and silicon Mach-Zehnder modulator operating at 25Gb/s,” *Optics Express*, vol. 24, no. 26, pp. 30 379–30 401, Dec. 2016.
- [73] M. A. Tran, D. Huang, J. Guo *et al.*, “Ring-Resonator Based Widely-Tunable Narrow-Linewidth Si/InP Integrated Lasers,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 2, pp. 1–14, Mar. 2020.
- [74] J. M. Ramirez, H. Elfaiki, T. Verolet *et al.*, “III-V-on-Silicon Integration: From Hybrid Devices to Heterogeneous Photonic Integrated Circuits,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 2, pp. 1–13, Mar. 2020.
- [75] “III-V material integration, IBM Research Zurich,” May 2018. [Online]. Available: <https://www.zurich.ibm.com/st/nanodevices/monolithicgrowth.html> (Accessed: 2021-08-30)
- [76] “Nanotechnology - Nanofabrication | Britannica.” [Online]. Available: <https://www.britannica.com/technology/nanotechnology/Nanofabrication> (Accessed: 2022-10-04)
- [77] J. G. Fiorenza, J.-S. Park, J. Hydrick *et al.*, “(Invited) Aspect Ratio Trapping: A Unique Technology for Integrating Ge and III-Vs with Silicon CMOS,” *ECS Transactions*, vol. 33, no. 6, pp. 963–976, Oct. 2010.
- [78] D. Van Thourhout, Y. Shi, M. Baryshnikova *et al.*, “Chapter Eight - Nano-ridge laser monolithically grown on (001) Si,” in *Semiconductors and Semimetals*, ser. Future Directions in Silicon Photonics, S. Lourdudoss, J. E. Bowers, and C. Jagadish, Eds. Elsevier, Jan. 2019, vol. 101, pp. 283–304. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0080878419300110> (Accessed: 2023-01-09)
- [79] H. Kim, W.-J. Lee, A. C. Farrell *et al.*, “Telecom-Wavelength Bottom-up Nanobeam Lasers on Silicon-on-Insulator,” *Nano Letters*, vol. 17, no. 9, pp. 5244–5250, Sep. 2017.
- [80] S. Mauthe, N. Vico Triviño, Y. Baumgartner *et al.*, “InP-on-Si Optically Pumped Microdisk Lasers via Monolithic Growth and Wafer Bonding,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 6, pp. 1–7, Nov. 2019.

- [81] D. Li, “Metalorganic Chemical Vapor Deposition (MOCVD),” in *Encyclopedia of Microfluidics and Nanofluidics*. New York, NY: Springer, 2015, pp. 1770–1770. [Online]. Available: https://doi.org/10.1007/978-1-4614-5491-5_200166 (Accessed: 2022-10-03)
- [82] P. Staudinger, S. Mauthe, N. V. Triviño *et al.*, “Wurtzite InP microdisks: From epitaxy to room-temperature lasing,” *Nanotechnology*, vol. 32, no. 7, p. 075605, Feb. 2021.
- [83] M. Kostrzewa, L. Di Cioccio, M. Zussy *et al.*, “InP dies transferred onto silicon substrate for optical interconnects application,” *Sensors and Actuators A: Physical*, vol. 125, no. 2, pp. 411–414, Jan. 2006.
- [84] D. Caimi, P. Tiwari, M. Sousa *et al.*, “Heterogeneous Integration of III–V Materials by Direct Wafer Bonding for High-Performance Electronics and Optoelectronics,” *IEEE Transactions on Electron Devices*, vol. 68, no. 7, pp. 3149–3156, Jul. 2021.
- [85] N. Daix, E. Uccelli, L. Czornomaz *et al.*, “Towards large size substrates for III-V co-integration made by direct wafer bonding on Si,” *APL Materials*, vol. 2, no. 8, p. 086104, Aug. 2014.
- [86] A. Higuera-Rodriguez, B. Romeira, S. Birindelli *et al.*, “Ultralow Surface Recombination Velocity in Passivated InGaAs/InP Nanopillars,” *Nano Letters*, vol. 17, no. 4, pp. 2627–2633, Apr. 2017.
- [87] H. Jansen, H. Gardeniers, M. de Boer *et al.*, “A survey on the reactive ion etching of silicon in microtechnology,” *Journal of Micromechanics and Microengineering*, vol. 6, no. 1, p. 14, Mar. 1996.
- [88] C. Cardinaud, M.-C. Peignon, and P.-Y. Tessier, “Plasma etching: Principles, mechanisms, application to micro- and nano-technologies,” *Applied Surface Science*, vol. 164, no. 1, pp. 72–83, Sep. 2000.
- [89] J. W. Coburn and H. F. Winters, “Ion- and electron-assisted gas-surface chemistry—An important effect in plasma etching,” *Journal of Applied Physics*, vol. 50, no. 5, pp. 3189–3196, May 1979.
- [90] P. Strasser, R. Wüest, F. Robin *et al.*, “Detailed analysis of the influence of an inductively coupled plasma reactive-ion etching process on the hole depth and shape of photonic crystals in InP/InGaAsP,” *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 25, no. 2, pp. 387–393, Mar. 2007.
- [91] K. Racka-Szmidt, B. Stonio, J. Żelazko *et al.*, “A Review: Inductively Coupled Plasma Reactive Ion Etching of Silicon Carbide,” *Materials*, vol. 15, no. 1, p. 123, Dec. 2021.
- [92] J. Lin, A. Leven, N. G. Weimann *et al.*, “Smooth and vertical-sidewall InP etching using Cl₂/N₂ inductively coupled plasma,” *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 22, no. 2, pp. 510–512, Mar. 2004.
- [93] A. Wieczorek, V. Djara, F. H. Peters *et al.*, “Inductively coupled plasma deep etching of InP/InGaAsP in Cl₂/CH₄/H₂ based chemistries with the electrode at 20 °C,” *Journal of Vacuum Science & Technology B*, vol. 30, no. 5, p. 051208, Sep. 2012.
- [94] H.-Y. Chen, H. E. Ruda, and A. Z. Navarro, “Inductively coupled plasma etching of InP using N₂/H₂,” *Journal of Applied Physics*, vol. 89, no. 10, pp. 5322–5325, May 2001.
- [95] S. Tachi, K. Tsujimoto, and S. Okudaira, “Low-temperature reactive ion etching and microwave plasma etching of silicon,” *Applied Physics Letters*, vol. 52, no. 8, pp. 616–618, Feb. 1988.

BIBLIOGRAPHY

- [96] “Binnig and Rohrer Nanotechnology Center, IBM Research Zurich,” May 2018. [Online]. Available: <https://www.zurich.ibm.com/brnc/> (Accessed: 2022-10-03)
- [97] C.-W. Lee and M.-K. Chin, “Room-Temperature Inductively Coupled Plasma Etching of InP Using Cl₂/N₂ and Cl₂/CH₄/H₂,” *Chinese Physics Letters*, vol. 23, no. 4, pp. 903–906, Mar. 2006.
- [98] C.-W. Lee, D. Nie, T. Mei *et al.*, “Study and optimization of room temperature inductively coupled plasma etching of InP using Cl₂/CH₄/H₂ and CH₄/H₂,” *Journal of Crystal Growth*, vol. 288, no. 1, pp. 213–216, Feb. 2006.
- [99] K. Vyas, K. M. Awan, and K. Dolgaleva, “Systematic study of InP/InGaAsP heated plasma etching and roughness improvement for integrated optical devices,” *Journal of Vacuum Science & Technology B*, vol. 41, no. 1, p. 012804, Jan. 2023.
- [100] P. Tiwari, A. Fischer, M. Scherrer *et al.*, “Single-Mode Emission in InP Microdisks on Si Using Au Antenna,” *ACS Photonics*, vol. 9, no. 4, pp. 1218–1225, Apr. 2022.
- [101] C.-Z. Sun, J.-B. Zhou, X. Bing *et al.*, “Vertical and Smooth, etching of InP by Cl₂/CH₄/Ar Inductively Coupled Plasma at Room Temperature,” *Chinese Physics Letters*, vol. 20, no. 8, pp. 1312–1314, Jul. 2003.
- [102] P. Tiwari, P. Wen, D. Caimi *et al.*, “Scaling of metal-clad InP nanodisk lasers: Optical performance and thermal effects,” *Optics Express*, vol. 29, no. 3, pp. 3915–3927, Feb. 2021.
- [103] H.-Y. Chen and H. E. Ruda, “Inductively coupled plasma etching of InP using CH₄/H₂ and CH₄/H₂/N₂,” *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 20, no. 1, pp. 47–52, Jan. 2002.
- [104] J. Antony, *Design of Experiments for Engineers and Scientists*, 2nd ed. Burlington: Elsevier Science, 2014.
- [105] Q. Li, C. Deeb, H. Debregeas *et al.*, “Room temperature inductively coupled plasma etching of InP with Cl₂ mixtures using SiO₂ and photoresist masks,” *Journal of Vacuum Science & Technology B*, vol. 42, no. 2, p. 022205, Feb. 2024.
- [106] A. Kuźmich, E. Pruszyńska-Karbownik, K. Pierściński *et al.*, “Mid-infrared quantum cascade laser waveguides with non-vertical sidewalls,” *Infrared Physics & Technology*, vol. 118, p. 103902, Nov. 2021.
- [107] J. E. Schramm, D. I. Babic, E. L. Hu *et al.*, “Fabrication of high-aspect-ratio InP-based vertical-cavity laser mirrors using CH₄/H₂/O₂/Ar reactive ion etching,” *J. Vac. Sci. Technol. B*, vol. 15, no. 6, 1997.
- [108] “The Open Educator - 1. Introduction to Design of Experiments.” [Online]. Available: <https://www.theopeneducator.com/doe/Introduction-to-Design-of-Experiments> (Accessed: 2022-11-22)
- [109] D. C. Montgomery, *Design and Analysis of Experiments*, ninth ed. Hoboken, NJ: Wiley, 2017.
- [110] C. Ebert, A. Bond, H. Cao *et al.*, “Selective area etching of InP with PCl₃ in MOVPE,” *Journal of Crystal Growth*, vol. 307, no. 1, pp. 92–96, Sep. 2007.

-
- [111] R. L. Plackett and J. P. Burman, “The Design of Optimum Multifactorial Experiments,” *Biometrika*, vol. 33, no. 4, pp. 305–325, Jun. 1946.
- [112] “pyDOE: The experimental design package for python — pyDOE 0.3.6 documentation.” [Online]. Available: <https://pythonhosted.org/pyDOE/index.html> (Accessed: 2023-01-11)
- [113] M. Fox and R. Ispasoiu, “Quantum Wells, Superlattices, and Band-Gap Engineering,” in *Springer Handbook of Electronic and Photonic Materials*, ser. Springer Handbooks, S. Kasap and P. Capper, Eds. Boston, MA: Springer US, 2007, pp. 1021–1040. [Online]. Available: https://doi.org/10.1007/978-0-387-29185-7_42 (Accessed: 2023-10-19)
- [114] X. Zhang, F. Zhang, R. Yi *et al.*, “Telecom-band multiwavelength vertical emitting quantum well nanowire laser arrays,” *Light: Science & Applications*, vol. 13, no. 1, p. 230, Sep. 2024.
- [115] Y. Arakawa and A. Yariv, “Quantum well lasers—Gain, spectra, dynamics,” *IEEE Journal of Quantum Electronics*, vol. 22, no. 9, pp. 1887–1899, Sep. 1986.
- [116] O. Svelto, *Principles of Lasers*, 5th ed. New York: Springer, 2010.
- [117] J. R. Hook, *Solid State Physics*, 2nd ed., ser. Manchester Physics Series. Hoboken: Wiley, 2013.
- [118] H. D. Young, *Sears and Zemansky’s University Physics: With Modern Physics*, 14th ed., ser. Always Learning. Harlow, England: Pearson, 2016.
- [119] O. Painter, R. K. Lee, A. Scherer *et al.*, “Two-Dimensional Photonic Band-Gap Defect Mode Laser,” *Science*, vol. 284, no. 5421, pp. 1819–1821, Jun. 1999.
- [120] M. Scherrer, C.-W. Lee, H. Schmid *et al.*, “Single-Mode Laser in the Telecom Range by Deterministic Amplification of the Topological Interface Mode,” *ACS Photonics*, vol. 11, no. 3, pp. 1006–1011, Mar. 2024.
- [121] A. Kodigala, T. Lepetit, Q. Gu *et al.*, “Lasing action from photonic bound states in continuum,” *Nature*, vol. 541, no. 7636, pp. 196–199, Jan. 2017.
- [122] S. L. McCall, A. F. J. Levi, R. E. Slusher *et al.*, “Whispering-gallery mode microdisk lasers,” *Applied Physics Letters*, vol. 60, no. 3, pp. 289–291, Jan. 1992.
- [123] L. Rayleigh, “CXII. The problem of the whispering gallery,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 20, no. 120, pp. 1001–1004, Dec. 1910.
- [124] “Ansys Lumerical FDTD | Simulation for Photonic Components.” [Online]. Available: <https://www.ansys.com/en-gb/products/optics/fdtd> (Accessed: 2025-04-08)
- [125] A. Jain, S. P. Ong, G. Hautier *et al.*, “Commentary: The Materials Project: A materials genome approach to accelerating materials innovation,” *APL Materials*, vol. 1, no. 1, p. 011002, Jul. 2013.
- [126] J. M. Munro, K. Latimer, M. K. Horton *et al.*, “An improved symmetry-based approach to reciprocal space path selection in band structure calculations,” *npj Computational Materials*, vol. 6, no. 1, pp. 1–6, Jul. 2020.
- [127] S. P. Ong, W. D. Richards, A. Jain *et al.*, “Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis,” *Computational Materials Science*, vol. 68, pp. 314–319, Feb. 2013.

BIBLIOGRAPHY

- [128] K. A. Johnson and N. W. Ashcroft, “Corrections to density-functional theory band gaps,” *Physical Review B*, vol. 58, no. 23, pp. 15 548–15 556, Dec. 1998.
- [129] A. Mishra, L. V. Titova, T. B. Hoang *et al.*, “Polarization and temperature dependence of photoluminescence from zincblende and wurtzite InP nanowires,” *Applied Physics Letters*, vol. 91, no. 26, p. 263104, Dec. 2007.
- [130] N. G. Basov, “Semiconductor Lasers,” *Science*, vol. 149, no. 3686, pp. 821–827, Aug. 1965.
- [131] M. G. A. Bernard and G. Duraffourg, “Laser Conditions in Semiconductors,” *physica status solidi (b)*, vol. 1, no. 7, pp. 699–703, 1961.
- [132] J. Y. Peng and X. B. Chen, “Modeling of Piezoelectric-Driven Stick–Slip Actuators,” *IEEE/ASME Transactions on Mechatronics*, vol. 16, no. 2, pp. 394–399, Apr. 2011.
- [133] I. D. W. Samuel, E. B. Namdas, and G. A. Turnbull, “How to recognize lasing,” *Nature Photonics*, vol. 3, no. 10, pp. 546–549, Oct. 2009.
- [134] W. E. Hayenga, H. Garcia-Gracia, H. Hodaie *et al.*, “Second-order coherence properties of metallic nanolasers,” *Optica*, vol. 3, no. 11, pp. 1187–1193, Nov. 2016.
- [135] N. M. Ravindra, P. Ganapathy, and J. Choi, “Energy gap–refractive index relations in semiconductors – An overview,” *Infrared Physics & Technology*, vol. 50, no. 1, pp. 21–29, Mar. 2007.
- [136] B. Bennett, R. Soref, and J. Del Alamo, “Carrier-induced change in refractive index of InP, GaAs and InGaAsP,” *IEEE Journal of Quantum Electronics*, vol. 26, no. 1, pp. 113–122, Jan. 1990.
- [137] T. L. Koch and J. E. Bowers, “Nature of wavelength chirping in directly modulated semiconductor lasers,” *Electronics Letters*, vol. 20, no. 25, pp. 1038–1040, Dec. 1984.
- [138] W. Shockley and W. T. Read, “Statistics of the Recombinations of Holes and Electrons,” *Physical Review*, vol. 87, no. 5, pp. 835–842, Sep. 1952.
- [139] S. Nakamura, “Background story of the invention of efficient blue InGaN light emitting diodes (Nobel Lecture),” *Annalen der Physik*, vol. 527, no. 5-6, pp. 335–349, 2015.
- [140] A. A. Ukhanov, A. Stintz, P. G. Eliseev *et al.*, “Comparison of the carrier induced refractive index, gain, and linewidth enhancement factor in quantum dot and quantum well lasers,” *Applied Physics Letters*, vol. 84, no. 7, pp. 1058–1060, Feb. 2004.
- [141] S. Melnik, G. Huyet, and A. V. Uskov, “The linewidth enhancement factor α of quantum dot semiconductor lasers,” *Optics Express*, vol. 14, no. 7, pp. 2950–2955, Apr. 2006.
- [142] E. Yablonovitch, B. J. Skromme, R. Bhat *et al.*, “Band bending, Fermi level pinning, and surface fixed charge on chemically prepared GaAs surfaces,” *Applied Physics Letters*, vol. 54, no. 6, pp. 555–557, Feb. 1989.
- [143] R. Iyer and D. L. Lile, “Role of polysulfides in the passivation of the InP surface,” *Applied Physics Letters*, vol. 59, no. 4, pp. 437–439, Jul. 1991.
- [144] V. N. Bessolov, M. V. Lebedev, and D. R. T. Zahn, “Surface of n-type InP (100) passivated in sulfide solutions,” *Semiconductors*, vol. 33, no. 4, pp. 416–420, Apr. 1999.
- [145] S. Mauthe, “Monolithic Integration of III-V Photonic Devices on Silicon,” Ph.D. dissertation, ETH Zurich, 2020. [Online]. Available: <https://www.research-collection.ethz.ch/handle/20.500.11850/443327> (Accessed: 2025-04-18)

-
- [146] L. K. van Vugt, S. J. Veen, E. P. A. M. Bakkers *et al.*, “Increase of the Photoluminescence Intensity of InP Nanowires by Photoassisted Surface Passivation,” *Journal of the American Chemical Society*, vol. 127, no. 35, pp. 12 357–12 362, Sep. 2005.
 - [147] A. Guivarc’h, H. L’Haridon, G. Pelous *et al.*, “Chemical cleaning of InP surfaces: Oxide composition and electrical properties,” *Journal of Applied Physics*, vol. 55, no. 4, pp. 1139–1148, Feb. 1984.
 - [148] M. Faur, M. Faur, P. Jenkins *et al.*, “Study of surface passivation of InP,” *Surface and Interface Analysis*, vol. 15, no. 12, pp. 745–750, 1990.
 - [149] Z. Xu, S. Zhang, J. Liang *et al.*, “Surface passivation of InGaN nanorods using H₃PO₄ treatment for enhanced photoelectrochemical performance,” *Journal of Power Sources*, vol. 419, pp. 65–71, Apr. 2019.
 - [150] M. L. Huang, Y. C. Chang, C. H. Chang *et al.*, “Surface passivation of III-V compound semiconductors using atomic-layer-deposition-grown Al₂O₃,” *Applied Physics Letters*, vol. 87, no. 25, p. 252104, Dec. 2005.
 - [151] P. Tiwari, “III-V on Si for Active Photonics: Downscaling and Emission Control by the Use of Metals,” Doctoral Thesis, ETH Zurich, 2021. [Online]. Available: <https://www.research-collection.ethz.ch/handle/20.500.11850/547800> (Accessed: 2025-04-18)
 - [152] P. Wen, P. Tiwari, M. Scherrer *et al.*, “Thermal Simulation and Experimental Analysis of Optically Pumped InP-on-Si Micro- and Nanocavity Lasers,” *ACS Photonics*, vol. 9, no. 4, pp. 1338–1348, Apr. 2022.
 - [153] L. Lin, Y. Xue, J. Li *et al.*, “C and L band room-temperature continuous-wave InP-based microdisk lasers grown on silicon,” *Optics Letters*, vol. 46, no. 12, p. 2836, Jun. 2021.
 - [154] Y. Wan, Q. Li, A. Y. Liu *et al.*, “Optically pumped 1.3 μ m room-temperature InAs quantum-dot micro-disk lasers directly grown on (001) silicon,” *Optics Letters*, vol. 41, no. 7, pp. 1664–1667, Apr. 2016.
 - [155] M. Athanasiou, R. Smith, B. Liu *et al.*, “Room temperature continuous-wave green lasing from an InGaN microdisk on silicon,” *Scientific Reports*, vol. 4, no. 1, p. 7250, Nov. 2014.
 - [156] V. Dolores-Calzadilla, B. Romeira, F. Pagliano *et al.*, “Waveguide-coupled nanopillar metal-cavity light-emitting diodes on silicon,” *Nature Communications*, vol. 8, no. 1, p. 14323, Feb. 2017.
 - [157] L. J. Hornbeck, “Multi-level digital micromirror device,” US Patent US5 583 688A, Dec., 1996. [Online]. Available: <https://patents.google.com/patent/US5583688A/en> (Accessed: 2025-04-22)
 - [158] D. Taillaert, F. Van Laere, M. Ayre *et al.*, “Grating Couplers for Coupling between Optical Fibers and Nanophotonic Waveguides,” *Japanese Journal of Applied Physics*, vol. 45, no. 8A, pp. 6071–6077, Aug. 2006.
 - [159] D. Saxena, A. Fischer, J. Dranczewski *et al.*, “Designed Semiconductor Network Random Lasers,” *Laser & Photonics Reviews*, vol. 19, no. 1, p. 2400623, 2025.
 - [160] “MODE - Finite Difference Eigenmode (FDE) solver introduction.” [Online]. Available: <https://optics.ansys.com/hc/en-us/articles/360034917233-MODE-Finite-Difference-Eigenmode-FDE-solver-introduction> (Accessed: 2025-04-08)

BIBLIOGRAPHY

- [161] Z. Zhu and T. G. Brown, “Full-vectorial finite-difference analysis of microstructured optical fibers,” *Optics Express*, vol. 10, no. 17, pp. 853–864, Aug. 2002.
- [162] D. Dai, “Advanced passive silicon photonic devices with asymmetric waveguide structures,” *Proceedings of the IEEE*, vol. 106, no. 12, pp. 2117–2143, 2018.
- [163] A. Arnaudon, D. Saxena, R. T. V *et al.*, “Arnaudon/netSALT: Paper release,” Zenodo, Sep. 2022. [Online]. Available: <https://zenodo.org/records/7092686> (Accessed: 2025-04-08)
- [164] L. Ge, Y. D. Chong, and A. D. Stone, “Steady-state ab initio laser theory: Generalizations and analytic results,” *Physical Review A*, vol. 82, no. 6, p. 063824, Dec. 2010.
- [165] M. Ahmed and M. Yamada, “Influence of instantaneous mode competition on the dynamics of semiconductor lasers,” *IEEE Journal of Quantum Electronics*, vol. 38, no. 6, pp. 682–693, Jun. 2002.
- [166] D. Smith, J. S. Myers, C. S. Kaplan *et al.*, “An aperiodic monotile,” *Combinatorial Theory*, vol. 4, no. 1, 2024.
- [167] X. Xu, M. Tan, B. Corcoran *et al.*, “11 TOPS photonic convolutional accelerator for optical neural networks,” *Nature*, vol. 589, no. 7840, pp. 44–51, Jan. 2021.
- [168] W. Schreiber, “Image processing for quality improvement,” *Proceedings of the IEEE*, vol. 66, no. 12, pp. 1640–1651, Dec. 1978.
- [169] “Cameraman,” 1978. [Online]. Available: <https://dome.mit.edu/handle/1721.3/195767> (Accessed: 2025-04-30)
- [170] C. Szegedy, W. Liu, Y. Jia *et al.*, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9. [Online]. Available: <https://ieeexplore.ieee.org/document/7298594> (Accessed: 2025-04-30)
- [171] M. Li, X. Li, W. Sun *et al.*, “Efficient convolutional neural network with multi-kernel enhancement features for real-time facial expression recognition,” *Journal of Real-Time Image Processing*, vol. 18, no. 6, pp. 2111–2122, Dec. 2021.
- [172] C. Szegedy, V. Vanhoucke, S. Ioffe *et al.*, “Rethinking the Inception Architecture for Computer Vision,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 2818–2826. [Online]. Available: <https://ieeexplore.ieee.org/document/7780677/> (Accessed: 2025-06-06)
- [173] X. Ding, X. Zhang, Y. Zhou *et al.*, “Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs,” *arXiv*, Apr. 2022. [Online]. Available: <http://arxiv.org/abs/2203.06717> (Accessed: 2025-04-30)
- [174] “MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges,” Dec. 2024. [Online]. Available: <https://web.archive.org/web/20241201065229/http://yann.lecun.com/exdb/mnist/> (Accessed: 2025-04-30)
- [175] Y. Lecun, L. Bottou, Y. Bengio *et al.*, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [176] W. H. Greene, *Econometric Analysis*, 7th ed. Harlow: Pearson Education, 2011.

-
- [177] N. Simon, J. Friedman, and T. Hastie, “A Blockwise Descent Algorithm for Group-penalized Multiresponse and Multinomial Regression,” *arXiv*, Nov. 2013. [Online]. Available: <http://arxiv.org/abs/1311.6529> (Accessed: 2025-05-06)
 - [178] S. W. Menard, *Applied Logistic Regression Analysis*, 2nd ed., ser. Quantitative Applications in the Social Sciences ; No. 07-106. Thousand Oaks, [Calif.] ;: SAGE, 2002.
 - [179] C. Zhu, R. H. Byrd, P. Lu *et al.*, “Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization,” *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, Dec. 1997.
 - [180] F. Pedregosa, G. Varoquaux, A. Gramfort *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
 - [181] T. Akiba, S. Sano, T. Yanase *et al.*, “Optuna: A Next-generation Hyperparameter Optimization Framework,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD ’19. New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 2623–2631. [Online]. Available: <https://doi.org/10.1145/3292500.3330701> (Accessed: 2025-05-06)
 - [182] “ZalandoResearch/fashion-mnist: A MNIST-like fashion product database. Benchmark.” [Online]. Available: <https://github.com/zalandoResearch/fashion-mnist> (Accessed: 2025-05-07)
 - [183] C. Peters, “Free blue noise textures,” Dec. 2016. [Online]. Available: <https://momentsingraphics.de/BlueNoise.html> (Accessed: 2025-05-07)
 - [184] R. Ulichney, “Dithering with blue noise,” *Proceedings of the IEEE*, vol. 76, no. 1, pp. 56–79, Jan. 1988.
 - [185] R. A. Ulichney, “Void-and-cluster method for dither array generation,” in *IS&T/SPIE’s Symposium on Electronic Imaging: Science and Technology*, J. P. Allebach and B. E. Rogowitz, Eds., San Jose, CA, Sep. 1993, pp. 332–343. [Online]. Available: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1009474> (Accessed: 2025-05-07)
 - [186] R. C. Deo, “Machine Learning in Medicine,” *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015.
 - [187] M. Shehab, L. Abualigah, Q. Shambour *et al.*, “Machine learning in medical applications: A review of state-of-the-art methods,” *Computers in Biology and Medicine*, vol. 145, p. 105458, Jun. 2022.
 - [188] M. L. Giger, “Machine Learning in Medical Imaging,” *Journal of the American College of Radiology*, vol. 15, no. 3, Part B, pp. 512–520, Mar. 2018.
 - [189] A. Mitani, A. Huang, S. Venugopalan *et al.*, “Detection of anaemia from retinal fundus images via deep learning,” *Nature Biomedical Engineering*, vol. 4, no. 1, pp. 18–27, Jan. 2020.
 - [190] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul. 2019.
 - [191] “Breast Histopathology 400X.” [Online]. Available: <https://www.kaggle.com/datasets/forderation/breakhis-400x> (Accessed: 2025-05-07)
 - [192] F. A. Spanhol, L. S. Oliveira, C. Petitjean *et al.*, “A Dataset for Breast Cancer Histopathological Image Classification,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016.

BIBLIOGRAPHY

- [193] L. Alzubaidi, J. Zhang, A. J. Humaidi *et al.*, “Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions,” *Journal of Big Data*, vol. 8, no. 1, p. 53, Mar. 2021.
- [194] M. R. Alom, F. A. Farid, M. A. Rahaman *et al.*, “An explainable AI-driven deep neural network for accurate breast cancer detection from histopathological and ultrasound images,” *Scientific Reports*, vol. 15, no. 1, p. 17531, May 2025.
- [195] I. Y. Tyukin, A. N. Gorban, M. H. Alkhudaydi *et al.*, “Demystification of Few-shot and One-shot Learning,” in *2021 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2021, pp. 1–7. [Online]. Available: <http://arxiv.org/abs/2104.12174> (Accessed: 2025-05-08)
- [196] D. Kerschensteiner, “Feature Detection by Retinal Ganglion Cells,” *Annual Review of Vision Science*, vol. 8, pp. 135–169, Sep. 2022.
- [197] R. H. Masland, “The neuronal organization of the retina,” *Neuron*, vol. 76, no. 2, pp. 266–280, Oct. 2012.
- [198] W. B. Thoreson and S. C. Mangel, “Lateral interactions in the outer retina,” *Progress in Retinal and Eye Research*, vol. 31, no. 5, pp. 407–441, Sep. 2012.
- [199] F. S. Werblin, “The retinal hypercircuit: A repeating synaptic interactive motif underlying visual function,” *The Journal of Physiology*, vol. 589, no. Pt 15, pp. 3691–3702, Aug. 2011.
- [200] A. L. Fairhall, C. A. Burlingame, R. Narasimhan *et al.*, “Selectivity for multiple stimulus features in retinal ganglion cells,” *Journal of Neurophysiology*, vol. 96, no. 5, pp. 2724–2738, Nov. 2006.
- [201] N. Perez-Nieves, V. C. H. Leung, P. L. Dragotti *et al.*, “Neural heterogeneity promotes robust learning,” *Nature Communications*, vol. 12, no. 1, p. 5791, Oct. 2021.
- [202] A. Krizhevsky and G. Hinton, “Learning multiple layers of features from tiny images,” Technical report, University of Toronto / University of Toronto, Toronto, Ontario, Tech. Rep. 0, 2009. [Online]. Available: <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf> (Accessed: 2025-05-21)
- [203] P. Tschandl, C. Rosendahl, and H. Kittler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Scientific Data*, vol. 5, no. 1, p. 180161, Aug. 2018.
- [204] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *arXiv*, Sep. 2020. [Online]. Available: <http://arxiv.org/abs/1905.11946> (Accessed: 2025-05-22)
- [205] K. Team, “Keras documentation: Image classification via fine-tuning with EfficientNet.” [Online]. Available: https://keras.io/examples/vision/image_classification_efficientnet_fine_tuning/ (Accessed: 2025-05-22)
- [206] W. K. Ng, “Functional control of nanophotonic complex lasers,” Ph.D. dissertation, Imperial College London, Jun. 2024. [Online]. Available: <http://hdl.handle.net/10044/1/115145> (Accessed: 2025-05-02)
- [207] P. Tschandl, C. Rinner, Z. Apalla *et al.*, “Human–computer collaboration for skin cancer recognition,” *Nature Medicine*, vol. 26, no. 8, pp. 1229–1234, Aug. 2020.
- [208] “HAM10000 Lesion Segmentations.” [Online]. Available: <https://www.kaggle.com/datasets/tschandl/ham10000-lesion-segmentations> (Accessed: 2025-05-22)