Center for Doctoral Training in Controlled Quantum Dynamics
Department of Physics, Imperial College of Science, Technology and Medicine

The role of reversibility in quantum thermodynamics and the foundations of quantum theory

Jonathan George Richens

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy of Philosophy in Physics of the University of London
October 2018

Abstract

The first half of this thesis is dedicated to the study of generalizations of probability theory which include quantum theory and many of its alternatives. This research program seeks to understand quantum theory by exploring it 'from the outside', where it sits within a landscape of all operationally defined theories. We approach this problem with a new perspective: generalized decoherence. Quantum theory recovers classical theory via decoherence, where a 'classical limit' is achieved by a quantum system interacting with its environment, losing its quantum coherence. Typically it is the emergence of the classical world from the quantum that is studied. We reverse this paradigm and ask - how does the existence of a classical limit define the structure of quantum theory, or any other post-classical theory? We derive the existence of entangled states as a necessary feature of any theory that allows for the emergence of the classical world in this way. We then use our framework for generalized decoherence to explore the properties of theories that can 'hyperdecohere' to quantum theory. We find that any such theory must be very exotic compared to quantum theory, violating either the postulate of tomographic locality, reversibility, causality or purity. In the second half of this thesis we turn our attention to the second law of thermodynamics. We present a framework for deriving the second law under general constraints and explore two pertinent examples, constraining the fluctuations in work and constraining the size of the thermal bath and deriving the second law in these cases. Bounding fluctuations results in a unified free energy that contains the single-shot and Helmholtz free energies as limiting cases. Bounding the size of the thermal bath results in a finite-bath second law that is more general that previous attempts and draws connections between non-asymptotic thermodynamics and second-order information theory.

List of publications

The work presented in this thesis contains material from the following publications

Al-Safi, Sabri W., and Jonathan Richens. "Reversibility and the structure of the local state space." New Journal of Physics 17.12 (2015): 123001.

Richens, Jonathan G., and Lluis Masanes. "Work extraction from quantum systems with bounded fluctuations in work." Nature communications 7 (2016): 13511.

Richens, Jonathan G., John H. Selby, and Sabri W. Al-Safi. "Entanglement is necessary for emergent classicality in all physical theories." Physical review letters 119.8 (2017): 080503.

Richens, Jonathan G., Ãlvaro M. Alhambra, and Lluis Masanes. "Finite-bath corrections to the second law of thermodynamics." Physical Review E 97.6 (2018): 062132.

Lee, Cairan M., John H. Selby and Richens, Jonathan G. "Quantum theory cannot be extended". In preparation

Acknowledgements

First and foremost I would like to thank my supervisor Lluis Masanes for taking me on and supporting me throughout my PhD. You're the smartest man I know, and in more ways than one I feel incredibly lucky to have been your student. I only write so little here because you taught me the value of brevity and simple, effective writing (which I renounce for the remainder of this thesis).

None of the work in this thesis would have happened without the gang of peers who patiently listened to my excited / half baked ideas and shared their own, better ideas. First there's John Selby, whose brilliant and lucid thinking, along with his mad squiggles, will surely lead to the next revolution in something or another. Thanks for teaching me to think and draw. Then there's Ciaran Lee who seamlessly combines the brains of a bona fide genius with the 'schlick' salesmanship of a future cult leader. Thanks for teaching me causality in a train station cafe Nero and for introducing me to the delights of Barry's tea. A special thanks to Sabri Al-Safi, a real researcher whose collaboration initiated my journey into quantum foundations research. Thanks for teaching me that proofs have to actually be true. I still don't quite believe it. Thanks to Alvaro Alhambra and Thomas Galley for being inspiring colleagues and offering support and chats over the years in the dungeon. Finally, thanks to Dan Goldwater for being a good friend and introducing me to anarchist sci-fi and sad disco. Ill be poorly imitating your beliefs and aesthetic for years to come.

A special thanks to Saurabh Johri for giving me a shot at doing something really exciting. I didn't think I would find anything more interesting than quantum physics - you proved me wrong. Thanks for taking a chance, and for overlooking my unorthodoxies as I adjust to civilized life.

For Liza, Pete and Eve, this thesis is for you. You are the three reasons I'm writing this in the first place. For Megan, this thesis is also for you. Thank you for keeping me sane and giving me inspiration when I needed it most.

Dedication

For Pete

Contents

Abstract **i**

Acknowledgements **v**

1 Introduction **1**

1.1 What is this thesis about? 1

1.2 Overview 7

1.3 Statement of Originality 9

1.4 Disclaimer 9

2 Generalized Probabilistic Theories **10**

2.1 Operational description of experiments 10

2.2 Probabilities from circuits 16

2.3 Tensor product composition 24

2.4 Causality 26

2.4.1 Recovering the convex sets formalism 32

2.4.2 Transformations 39

2.4.3 Composition 43

3 Generalized Post-Classical Theories 49

3.1 Decoherence 49

3.2 Decoherence in theories without entanglement 58

3.3 Deriving the existence of entangled states 59

3.3.1 proof of the main result 68

3.3.2 Single-system consequences of decoherence 68

3.3.3 Composite system consequences of decoherence 71

3.3.4 Proof of main theorem 74

3.3.5 Discussion of results 76

4 Hyperdecoherence 78

4.0.1 Existing no-go theorems 79

4.0.2 No-go theorem on hyper-decoherence 84

4.1 Discussion 88

4.2 Conclusions 93

5 Thermodynamics of individual quantum systems 95

5.0.1 Thermal baths 95

5.0.2 Thermal operations 98

5.0.3 Thermal operations with non-constant Hamiltonians 108

5.0.4 Work, heat and transformations 111

5.0.5 Landauer erasure: a simple illustration . . . 114

5.1 Fluctuating work . . . 115

5.2 Deterministic work extraction . . . 121

5.3 Towards a unified free energy . . . 123

5.4 Finite bath corrections to the second law . . . 139

5.4.1 Finite bath corrections to the second law . . . 141

5.5 Conclusions . . . 145

Appendices **147**

.1 Thermal operations . . . 148

.1.1 determining the optimal map . . . 148

.1.2 Finding the optimal partition . . . 154

.1.3 Optimal work of formation . . . 160

.1.4 Optimal partitions . . . 164

.1.5 Asymptotics . . . 165

.1.6 Fluctuation bounds . . . 167

.1.7 c-bounded work and the Jarzinski equality . . . 169

.2 *c*-bounded Carnot efficiency . . . 171

.3 Finite bath second law . . . 175

.3.1 finite bath thermal operations . . . 176

.4 Permissions . . . 183

Bibliography **183**

List of Figures

2.1 Diagram depicts a particular run of an experiment. The classical pointers are shown as coloured lights on the devices. The experiment involves two devices that prepare systems A_1 and A_2 in one of two possible states, a device that then performs one of three possible transformations on these systems and then outputs systems B_1 and B_2, and a final device that performs a joint measurement on these systems, resulting in the outcome 'red'. 11

2.2 Convex set depicting a state space Ω_A. ρ_1 and ρ_2 are extremal points in the set - pure states. Note that ρ_1 is exposed, as it lies on the supporting hyperplane given by $e(x) = 1, x \in \Omega_A$, however ρ_2 is unexposed (there is no hyperplane that supports it, it sits between a continuous curve and an edge). The minimal face containing ρ_1 and ρ_3 is the edge $e(x) = 1$. State $\sigma = p\rho_1 + (1-p)\rho_2$, but also note that ρ_1, ρ_2, ρ_3 all refine σ (as indeed do all states in the convex set, as σ is internal). 36

2.3 Qutrit subspace 39

3.1 Taking the local state space of the PR-box / gbit and composing it with a 0-dimensional state space under the direct sum results in a square-based pyramid state space. Its decomposability (note that by construction $\mathsf{span}K_1 \cap \mathsf{span}K_2 = 0$) results in the existence of a 'non-disturbing measurement'. This measurement has effects e and its complementary effect $u - e$ which deterministically distinguish states on the square base with the state atthe point of the pyramid. It is possible to measure if an unkown state is in Ω_1 or Ω_2 without disturbing it. However, Ω_1 generates an indecomposable sub-cone, and thus there can be no consistent measurement that can give information on an unknown state in Ω_1 without disturbing it. As shown in [4], there are reversible interactions that correlate two such systems without altering the state of one of them, allowing non-disturbing measurements to be dynamically constructed as in Definition 21 . . . 64

3.2 The square state space (the local state space of PR-boxes [110]) has a face lattice of extremal states ρ_i, $i = 1, \ldots, 4$, and edges which are 'picked out' by local extremal effects $\mathcal{E}_{ij}(\rho) = 0$. 66

5.1 figure from [95], reprinted with permission. 114

5.2 How c-bounded work extraction and state formation varies further from equilibrium. Figure shows the unbounded work $W(\rho)$, the c-bounded work content $W^{(c)}(\rho)$ and the c-bounded work of formation $W_F^{(c)}(\rho)$ for the state $\rho = p\,|0\rangle\langle 0| + (1-p)\,|1\rangle\langle 1|$ with Hamiltonian $H_S = E\,|0\rangle\langle 0|$ with $\beta = 1$, $E = 0.1$ and $c = 0.7$. In the above figure x denotes the probability $1 - p$ of the system being in the excited state $|1\rangle$. Note that despite choosing a c-bound that allows for fluctuations of the order of the maximal work that can be extracted from the pure state $p = 1$, in general we can extract much less that this amount. There is a discontinuity in $W^{(c)}(\rho)$ at $x = 0$ where we recover $W^{(c)}(\rho) = W^{(\infty)}(\rho)$. Notice also that closer to the thermal state the dissipation (difference between the $W^{(c)}(\rho)$ or $W_F^{(c)}(\rho)$ and $W^{(\infty)}(\rho)$) is greater for state formations than work extraction, and this reverses as the state moves further from the Gibbs state. 133

5.3 How c-bounded work extraction and state formation varies with c for a fixed initial / final state. figure shows $W_F^{(c)}(\rho)$ and $W^{(c)}(\rho)$ v.s. c for the trit state $\rho = 0.7\,|0\rangle\langle 0| + 0.2\,|1\rangle\langle 1| + 0.1\,|2\rangle\langle 2|$ with Hamiltonian $H_s = E_1\,|0\rangle\langle 0| + E_2\,|1\rangle\langle 1|$ with $E_1 = 0.1$, $E_2 = 0.2$. β is set to 1. x denotes the value of c, which we tune between 0 and 1.5. For small c the dissipation $|W^{(c)}(\rho) - W^{(\infty)}(\rho)|$ for the formation protocol is greater than for the extraction protocol, and for large c the relationship is inverted. Note that for $c > 0.9$ it possible to thermodynamically reversibly prepare state ρ but not to thermodynamically reversibly extract work from it. 134

5.4 The corrected Carnot efficiency for various values of c for a single qubit heat engine. Figure shows the c-bounded Carnot efficiencies vs. T_H for single qubit engine with gap $E = 0.1$ and $T_C = 1$. The black line gives the unbounded Carnot efficiency. The Blue lines give the c-bounded efficiencies with there corresponding c's marked on the figure. The dashed lines give the maximum attainable efficiency in the limit of asymptotic temperature difference between $\beta_C/\beta_H \to \infty$. 137

5 Lorenz curve depicting work of formation of a state. Figure shows the thermally sharp state whose single-shot work content gives the work of formation of state ρ'. This also represents the maximal average work that can be extracted from ρ' 167

Chapter 1

Introduction

1.1 What is this thesis about?

This is a thesis of two halves. In the first half we study the physical laws that give rise to the structure of quantum theory. Since the early 20^{th} century we have known that nature is non-classical, but we have not known exactly *why* this should be the case, and why it should specifically fall to quantum theory to explain this non-classicality. The work presented in these first chapters was initially motivated by a simple question: could a pre-quantum physicist, armed with only the knowledge that the world is non-classical and some basic physical principles, derive part or all of quantum theory? Or to put it simply, is quantum theory *inevitable* given some set of physical principles?

To answer this question, we construct a new class of *post-classical theories* - theories that can recover classical theory through a process analogous to dynamical decoherence. We show that one of the most non-classical features of quantum theory, entanglement, is an inevitable feature of all post-classical theories and hence is required for the emergence of the classical world. We then extend our framework to characterize *post-quantum* theories - theories that can recover a quantum limit through 'hyperdecoherence'. We use this approach to derive the strongest

constraints to date on what these theories can look like. Ultimately we show that any post-quantum theory of this form must violate at least one of four well motivated physical postulates that are obeyed by quantum theory; reversibility, local tomography, causality or purity.

Why is this interesting? Our first result, the inevitability of entanglement, would allow a physicist with no knowledge of quantum theory to derive the existence of entangled states, armed only with the knowledge that nature is non-classical. This constitutes the first derivation of entanglement from a minimal and physical set of principles. Similarly our second result, a no-go theorem on post-quantum theories, allows a physicist to derive some of the 'odd' features of any possible post-quantum theory, just from the assumption that quantum theory breaks down at some level. This gives us a principled way in which to speculate about what post quantum theories could look like.

In the second half of this thesis we turn our attention to the foundations of quantum thermodynamics. Historically, the laws of thermodynamics have proven very successful at describing macroscopic thermodynamic protocols and systems like heat engines and thermal baths. However, as we start to construct these systems at the mesoscopic and microscopic level, we can no longer make the same assumptions that we make when describing large-scale, classical thermal systems. Not only do quantum effects come into play in these new regimes, but classical phenomena such as work fluctuations and finite size/time effects can no longer be ignored. To build a new framework for describing thermodynamics in this new 'small scale' regime, we need to take these effects seriously and to clarify what the key theoretical concepts like heat, work and entropy really mean in this regime [6, 130, 108, 2].

We focus on the second law of thermodynamics, which encodes the ultimate limitations on the thermal operations we can achieve with a system as it interacts with a thermal environment. We show how the standard second law can be derived from the postulates of energy conservation and microscopic reversibility alone, and

present a new framework for deriving the second law under arbitrary additional constraints. We then derive corrections to the second law under two constraints that are very pertinent to small scale systems, bounding the size of fluctuations in work and bounding the size of the thermal bath. First we derive a new expression for the work that can be extracted from a quantum system when we bound the size of work fluctuations. We then turn our attention to removing the standard assumption that the thermal bath is infinite in size, and derive bounds on the optimal work that can be achieved using a thermal bath with a finite volume / time interaction with a bath.

Why is this interesting? There has been a great deal of work recently on how we define work for individual quantum systems, which has resulted in two distinct approaches. In the first approach, work fluctuations are allowed to be arbitrarily large [124] whereas in the second approach, work is constrained to be a deterministic quantity [2, 71]. The 'bounded-fluctuation free energy' that we derive contains these two approaches as special limiting cases within a single unified framework. From a practical perspective, our results can be applied directly to the ongoing development of small scale heat engines and refrigerators [89, 123, 49, 117, 31] which due to their size are inherently fragile and cannot accommodate arbitrarily large work fluctuations. Moving on to the finite-bath second law, our results present a universally applicable upper bound to the work extracted by a protocol operating in finite space and / or finite time. This is particularly relevant to systems like small scale heat engines that operate over very small time scales and interact with small regions of their environments.

Before introducing the structure of this thesis, it will be useful to discuss some of these statements with a little more depth. First we clarify the goals of the first section of this thesis, namely; what we mean by physical postulates and laws, how we go about finding them, and what our results mean in a broader context.

In the first section of this thesis we work within the framework of generalized prob-

abilistic theories [15, 67, 37]. This framework encapsulates all possible theories of nature that can describe the probabilities of experimental outcomes, and allows us to study the features of quantum theory 'from the outside', i.e. from within the space of all possible theories. To understand why it is useful to consider quantum theory from this perspective, it is first necessary to consider the current formalism of quantum theory.

Quantum theory emerged gradually over the last century, and the current dominant formalism was penned by Von Nuemann in "Mathematische Grundlagen der Quantenmechanik" over 80 years ago [101]. It is difficult to see any resemblance between the pleasing formalism of Von Nuemann's quantum theory and its messy predecessor, matrix mechanics [22]. Although the formalism has progressed, the physical understanding has (until recent years) stayed put. A nice formalism is poor substitute for understanding. If you ask someone to explain what quantum theory means and they start talking about Hilbert spaces and C^* algebras then, while they can do quantum theory, I don't consider this to be a satisfactory understanding of quantum theory. Von Nuemann quantum theory is no more explanatory than matrix mechanics. One of the goals of quantum foundations research is to reconstruct quantum theory from within an operation framework, such that its fundamental physical principles become evident [67, 14, 33].

Why should it be useful to consider quantum theory from within a different formalism? Firstly, this relates to the goal of deriving the physical principles behind quantum theory. A physical principle is a statement of the constraints obeyed by a theory that makes reference only to *operational primitives*, such as the possible states, transformations and observations that can be achieved when operating within that theory. Special and general relativity are good examples of theories that were built from physical principles. These theories were originally operationally defined theories of clocks and rigid rods, from which a mathematical formalism was later developed. For example, the no-signalling principle and the equivalence principle are operational constraints with a clear physical interpretation. We cannot hope

to truly understand quantum theory without knowing how to formulate it in terms of physical principles, no more than we could understand special relativity without knowing that the speed of light is absolute.

But do we really need to understand quantum theory in order to develop it, and perhaps one day go beyond it? There is actually a good historical justification for believing that we do not need this deeper understanding in order to make progress. Most physical laws have been constructed without a preexisting understanding of their underlying assumptions. For example, almost every physical law makes use of the notion of causality at some level, whilst the precise mathematical and operational statement of 'A causes B' has only existed for 30 years [106]. Newton's theory of gravity does not rely on Pearl's do-calculus. However, this doesn't mean we can continue to ignore foundational questions and expect to make progress. We can develop causal laws because we have an innate understanding of classical causality pre-baked into our cognitive models of the world [127, 55]. With the advent of quantum physics we have entered a new territory where our classical notions no longer apply. This is an unprecedented situation in the history of science; we have lost the privilege of resting on our intuitions.

The second benefit of reformulating quantum theory is that different formulations give us different perspectives on the same object. This is invaluable if we try to go beyond quantum theory, e.g. to alter one or more of its postulates. Naive attempts to extend upon quantum theory from the perspective of its mathematical postulates have failed by violating very fundamental physical laws. For example, relaxing the postulate of unitarity to give non-linear quantum theory leads to violations of the no-signalling principle [1]. Be reformulating quantum theory in terms of different physical postulates we can relax its structure in different (and hopefully more fruitful) directions. In the latter part of the second half of this thesis we make some steps towards achieving this goal, by deriving a set of physical principles obeyed by quantum theory that must be relaxed in order to construct viable post-quantum theories.

Where is this part of the thesis placed within a broader context? There has been significant work on understanding the mathematical objects of quantum theory - primarily focused on how we should interpret the quantum wavefunction (see [29] for a survey on the consensus of interpretations). We refer to this body of work as *interpretational* foundations. We focus on a different approach which involves looking for these aforementioned physical postulates. Put simply, we don't care what a wavefunction is, so long as we can say 'this plus this plus this equals quantum theory', and where every 'this' is a physical principle. There has been great progress in this area, mostly occurring in the past 20 years [65, 67, 95, 33, 45, 116, 70, 69, 34, 120, 14], but a consensus on the correct physical postulates of quantum theory has yet to be reached.

Most of this progress has been focused on deriving the entirety of quantum theory from some physical postulates [14, 33, 120]. In this thesis we take a different tack. If quantum theory is not the bedrock of theory-space, then it is likely that some of its features carry over to whatever supersedes it. There is then an advantage in considering its features in isolation. By features, we mean the objects of physical statements, like entanglement, coherence, entropy, and so on. Furthermore, we consider how these features emerge *between* theories. For example, if quantum theory is a super-theory of classical theory (classical theory is a sub-theory of quantum theory), what does this tell us about its physical features?

The first half of this thesis sets out to begin this line of research. We give rigorous and general definitions of sub-theories and super-theories and the processes that map between them. This framework is then used to derive our main results. To date, the only well studied super-theory / sub-theory relation is that of quantum theory to classical theory, but there could be a rich hierarchy and structure of generalized probabilistic super and sub theories waiting to be explored using the methods outlined in this thesis.

How do these two parts, thermodynamics and quantum foundations, thread to-

gether? There are some physical postulates that appear again and again in a variety of physical laws. The central theme of this thesis will be one of these postulates - reversibility. The postulate states that when a system evolves from one state to another, it is (at least in principle) possible to return it to its original state. This seemingly innocuous assumption imparts very strong constraints on the allowed dynamics and kinematics. As we will show, quantum theory and quantum and classical thermodynamics are completely contingent on the assumption of reversibility. For example, we show that if we can violate reversibility then we can systematically violate the second law of thermodynamics. In the first section, reversibility is a necessary assumption in all of our proofs. In the second section, we use the postulate of reversibility as the primary constraint in deriving the second law of thermodynamics and the second law with corrections.

We now outline the content of this thesis.

1.2 Overview

- Chapter 2. In this chapter we outline the framework of generalized probabilistic theories (GPTs), which will be used to derive the results in chapters 2 - 4. This framework is broad enough to describe any theory that can assign probabilities to the outcome of any currently conceivable experiment. We detail the fundamental processes (preparations, measurements and transformation), and describe how these can be composed in series and parallel to form experiments. We start with a diagrammatic, operational description of experiments (process theories) and use this to derive the convex sets formalism of GPTs. We introduce the key physical principles that we consider in the following chapters, namely tomographic locality and causality.

- Chapter 3. In this chapter we present our framework for describing the process of decoherence in general theories. We use this to define the notions of sub

and super theories, and define a new class of *post-classical* theories; theories that can recover classical probability theory through the process of non-trivial decoherence. We then use these results to derive the necessity for the existence of entangled states in all post-classical theories. To do so we define the class of 'non-disturbing' measurements and show how the existence of these measurements constrain systems to be classical. This chapter contains the results from [115] and [4] and the results are derived using the convex sets formalism of GPTs.

- Chapter 4. In this chapter we extend the framework presented in chapter 4 to consider the possibility of *hyperdecoherence*. That is, the possibility of the existence of post-quantum theories that recover quantum theory through processes analogous to dynamical decoherence. We review current results along these lines [86], and outline how we will remove some of their assumptions. We then prove that any theory that hyperdecoheres to quantum theory must its self violate either the postulates of causality, reversibility, purity of the sub-theory, or tomographic locality. We finish this chapter of the thesis by briefly discussing these postulates in depth and what it could mean to try to construct a theory without them.

- Chapter 5. In this chapter we turn our attention to the foundations of quantum thermodynamics. We introduce the key components of our framework for describing heat baths, work extraction and the second law of thermodynamics under general constraints. We briefly discuss the current standard work extraction protocols, namely single-shot and standard (Helmholtz) work extraction. We then present our case for constructing a single unified free energy that encapsulates these two formalisms as limiting cases. We provide a derivation of our unified free energy. Finally, we use our formalism to remove the assumption of infinite heat baths. We present a general definition of large but finite baths, and derive a corrected second law under these assumptions.

1.3 Statement of Originality

Except when otherwise acknowledged and referenced within the text, the following thesis is my own original work that took place during my PhD at Imperial College. It has not been already submitted to satisfy any degree requirement at this or any other university. The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Jonathan Richens

November 2018

1.4 Disclaimer

In this thesis I present the results of the following papers, first-authored by myself [113, 114, 115]. In some instances I will quote results directly from these papers. I do not include any results that are not my original work unless otherwise stated, and any direct quotations are cited to their original source. In some places passages from my own publications, [113, 114, 115], have been quoted verbatim. In these few cases the original text was written by myself and has been re-printed with permission by the publishing journals.

Chapter 2

Generalized Probabilistic Theories

In this chapter we present the framework of *generalized probabilistic theories* (GPTs). This is the framework within which we will derive all results in Chapters 2-4. We introduce this framework diagrammatically at first, as this gives the most intuitive explanation for the more standard convex sets formalism of GPTs, which we then use to derive our results in Chapter 3. The diagrammatic framework we employ, which follows that laid out in [33, 32], allows for some proofs to be greatly simplified and we employ it extensively in Chapter 4.

2.1 Operational description of experiments

Intuitively, an experiment is described by a collection of experimental devices that are connected to each other in series and parallel. In a run of the experiment the systems that we are conducting the experiment on (a photon, an atom, etc) move between these devices in a fixed causal direction (forwards in time). The experimenter can intervene to change the setting of the devices, and can read out any information returned by the devices, which we associate with *classical pointers* on the device. For example, the output of a device could be 'detector clicks', indicating the presence of a particle being detected in some position. In general we expect

that over many runs of an experiment the output to vary, as described by some probability distribution.

For example, one such experiment is depicted in Figure 2.1. A preparation device prepares a pair of systems independently, in one of two states with some probability. The preparation device indicates which state was prepared by switching on a green light or a red light. These two systems then undergo some transformation, for example some interaction takes places between them with some probability. This system is then passed to a device that performs a two-outcome measurement on the systems, indicating the outcome of that measurement by switching on a green or red light.

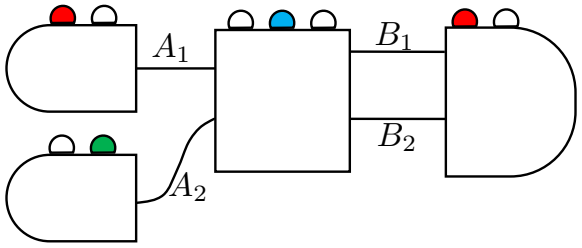


Figure 2.1: Diagram depicts a particular run of an experiment. The classical pointers are shown as coloured lights on the devices. The experiment involves two devices that prepare systems A_1 and A_2 in one of two possible states, a device that then performs one of three possible transformations on these systems and then outputs systems B_1 and B_2, and a final device that performs a joint measurement on these systems, resulting in the outcome 'red'.

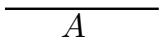
As we will show, even this seemingly innocuous description of what an experiment can look like imparts strong constraints on the type of theory that can describe the experiment and, conversely, the type of theory any such experiment can test. For example, we are tacitly making the assumption of *causality* (which we introduce later).

Our goal is to arrive at a rigorous framework for describing experiments like Figure 2.1 regardless of the specific laws that describe the systems and devices we have access to. That is, we desire a *theory-independent* framework for describing experimental outcomes. This framework must be capable of capturing the predictions of classical probability theory, quantum theory, and any other self-consistent the-

ory that can predict the outcomes of experiments. The framework should present a minimal and mathematically rigorous description of these experimental devices, how they can be composed, and how we then infer the probabilities of experimental outcomes. This framework, which we now present, is precisely the 'generalized probabilistic theories' (GPTs) framework.

In the following, we present a derivation of the key components of the GPT framework from diagrammatic definitions. This allows us to arrive at the mathematical constraints imposed on GPT's in a clear and logical way, rather than simply stating mathematical assumptions. For reference, the approach to GPT's presented here most closely resembles that given in [34, 32, 33].

Firstly we introduce the primitive notion of a physical system, which for now we denote with Roman capitals $A, B, \ldots$. The experimental devices act on these physical systems, and the path of these systems between devices are depicted by wires labeled by the system type



(2.1)

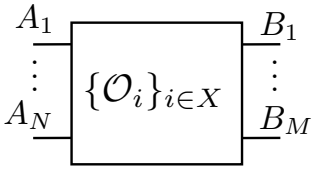
Although it is possible for devices to change the type of a system, for example a device that takes in a system of type A, discards it, and prepares a system of type B, we do not allow wires of different types to be directly connected, as the type of a system is preserved unless some process acts upon it such as to change it's type.

We refer to a particular device as a *process*. In general a process can result in multiple different *events*, such as a preparation device that can prepare a given system in by sampling from some ensemble of possible states. We label the set of possible events associated with a process by a classical pointer X. We assume that the pointer of a given process has a finite number of configurations, $i \in X$ where X

is a finite set. Each value of the pointer corresponds to a different *event* associated with this process, namely preparing a specific state, performing a transformation or returning a measurement outcome.

Definition 1 (Processes)**.** A process $\mathcal{O}$ denotes the single use of a piece of experimental apparatus with input systems labeled $A_1, A_2, \ldots, A_N$ and output systems labeled $B_1, B_2, \ldots, B_M$. When the device is used it produces outcome i from a possible finite number of events in a finite outcome set X, with each operation represented by the event operator $\mathcal{O}_i$. We often denote a process by the associated set of events $\{\mathcal{O}_i\}_{i\in X}$.

Diagrammatically, we represent such a process as as

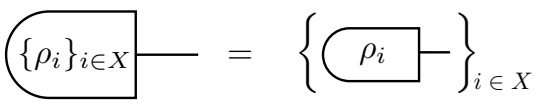


(2.2)

An important class of processes are those that prepare a system, or collection of systems, in some initial state. We call these processes *preparations*.

Definition 2 (Preparations)**.** A preparation ρ is a process with no inputs and one or more outputs. The associated set of events $\{\rho_i\}_{i\in X}$ correspond to a specific state prepared in a giving run of the experiment.

Diagrammatically, we represent preparations as 'cups'. As the $\{\mathcal{O}_i\}_{i\in X}$ represent different state preparations, we drop the calligraphic notation and represent states with Greek letters



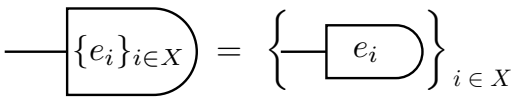
(2.3)

For example, the preparation $\{\rho_i\}_{i\in X}$ represents an apparatus that prepares a system in a state belonging to the collection of states indexed by $i \in X$, along with some probability distribution over this ensemble.

Measurements are described by processes that have one or more wires in and no wires out. That is, we can represent a measurement on system A as a 'black box' process that takes in system A and returns some classical probability distribution, typically discarding the system in the process[1]. The events associated with measurements are referred to as *effects*, where an effect specifies a particular outcome of a measurement process.

Definition 3 (Measurements)**.** A measurement $\mathcal{E}$ is a process with one or more inputs and no outputs. The associated set of events correspond to a specific measurement outcome and are called *effects* $\{e_i\}_{i\in X}$.

We depict measurements and effects as 'caps'



(2.4)

Processes with both inputs and outputs are referred to as *transformations*, which we denote with calligraphic capitals $\mathcal{T}$. Using the notation of [32], we denote the set of states on system A as $\mathsf{St}(A)$, the set of effects on system A as $\mathsf{Eff}(A)$ and the set of transformations from A to B as $\mathsf{Trans}(A, B)$.

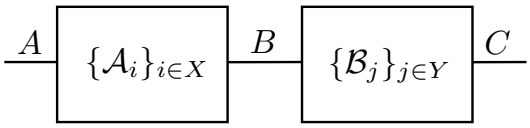
Now that we have defined the necessary processes to describe an experiment, we need some way of composing them together to form experimental circuits. Composition of two processes in series can be achieved between two processes if the outputs of one have the same system type as an inputs of the proceeding process. For example

[1] Although, as we will see, there are important classes of measurements that do not discard their input systems

consider the process $\{\mathcal{A}_i\}_{i\in X}$ that takes in systems of type A and outputs systems of type B, and process $\{\mathcal{B}_j\}_{j\in Y}$ that takes in systems of type B and outputs systems of type C. These can be composed with process $\mathcal{A}$ proceeding $\mathcal{B}$ (but not vice-versa unless $C = A$). Algebraically, we denote this composition as

$$\{\mathcal{A}_i\}_{i\in X} \circ \{\mathcal{B}_j\}_{j\in Y} = \{\mathcal{A}_i \circ \mathcal{B}_j\}_{(i,j)\in X\times Y} \tag{2.5}$$

e.g. the composed processes now form a process from A to C with events indexed by the product set $(i, j) \in X \times Y$. In our diagrams we chose by convention for time to flow from left to right (or from bottom to top), and as such the series composition of process $\mathcal{A}$ followed by process $\mathcal{B}$ is depicted as

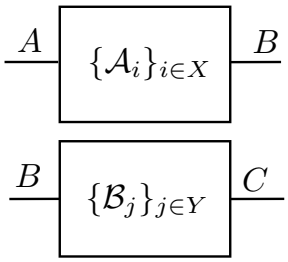


(2.6)

Note that as 'time' flows from left to right in circuit diagrams, the series composition of processes never admit causal loops (acyclicity). If we composed processes $\mathcal{A}$ and $\mathcal{B}$ in parallel, the result would be a process acting on the composite system AB and outputting a composite system of type BC. We denote this parallel composition with the symbol $\otimes$.

$$\{\mathcal{A}_i\}_{i\in X} \otimes \{\mathcal{B}_j\}_{j\in Y} = \{\mathcal{A}_i \otimes \mathcal{B}_j\}_{(i,j)\in X\times Y} \tag{2.7}$$

which is depicted simply as



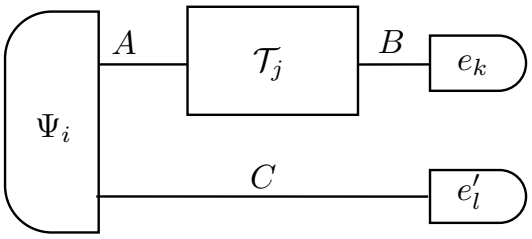
(2.8)

Note that as it stands $\otimes$ only represents the diagrammatic operation of parallel composition of processes at this point. In general $\otimes$ need not correspond to the standard vector tensor product, although we will introduce additional postulates that constrain it to be this. Diagrammatically, it is trivial to show that under these laws of composition, series and parallel composition operators are associative

$$(\mathcal{A}_i \circ \mathcal{C}_k) \otimes (\mathcal{B}_j \circ \mathcal{D}_l) = (\mathcal{A}_i \otimes \mathcal{B}_j) \circ (\mathcal{C}_k \otimes \mathcal{D}_l) \tag{2.9}$$

2.2 Probabilities from circuits

For these processes to meaningfully represent experiments we need to be able to represent the probabilities of any event that can happen in the experiment. This requires that we introduce a final object, the *scalar*, which is a process with no input or output wires. One example of such a process is a closed circuit with no free inputs or outputs remaining, which is a scalar that corresponds to the probability of the events depicting in the circuit having occurred in a run of the experiment, for example the circuit



$$\sim \; P(k,\, l,\, i,\, j) \tag{2.10}$$

is a closed diagram of events, representing the scalar $P(k,\, l,\, i,\, j)$ which is the joint probability of the events $i,\, j\; k,\, l$ occurring in a run of the experiment.

The correspondence between closed circuits and probabilities imparts additional constraints on the operational process theory described in the previous section. This correspondence between closed circuits and probabilities can be summarized by the assumption that any operational theory should be capable of predicting the outcome of experiments, and we refer to operational theories as *generalized probabilistic theories*.

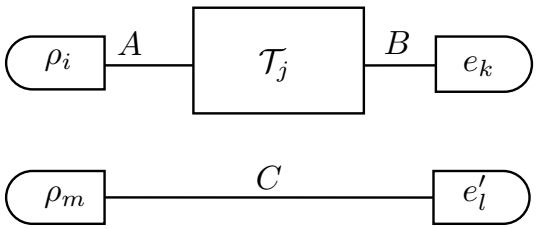
Definition 4 (Generalized probabilistic theory)**.** A generalized probabilistic theory is a set of processes $\{\mathcal{O}_i\}$ that are closed under series $\circ$ and parallel $\otimes$ composition, such that each closed circuit containing processes $\{\mathcal{A}_i\}_{i\in X}$, $\{\mathcal{B}_j\}_{j\in Y}$, $\{\mathcal{C}_k\}_{k\in Z}, \ldots$ corresponds to a set of probabilities $\{p_a\}_{a\in X\times Y\times Z\times\ldots}$ s.t. $p_a \in [0,1]\; \forall\; a$ and obeying the law of total probability $\sum_{a\in X\times Y\times Z\times\ldots} p_a = 1$.

Note that by Definition 4, for a closed circuit that separates into two or more independent closed circuits, the probabilities associated with these independent circuits factorize, as the sets of event indexes factorize over the two circuits.

$$X \times Y \times Z \times \cdots = X_1 \times Y_1 \times Z_1 \times \ldots, X_2 \times Y_2 \times Z_2 \times \ldots \tag{2.11}$$

For example, in the circuit below depicting single events for all processes the joint probability for the overall circuit factories into independent probabilities for the two

independent circuits.



$$\sim\ P(i,\, j,\, k) \times P'(m,\, l) \tag{2.12}$$

The introduction of probabilities imparts a linear structure on operational theories which, as we will now show, allows us to construct a real vector space representation for processes. In the following we use the 'Dirac'-type notation to denote states $|\rho_i)_A \in \mathsf{St}(A)$ and effects $(e_j| \in \mathsf{Eff}(B)$ and denote the closed circuit between a state and effect as

$$(e_j|\rho_i) = P(i,\, j) \tag{2.13}$$

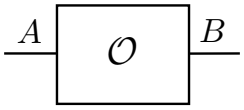
which is clearly linear as $(e_j|\,(\alpha\,|\rho_i) + \beta\,|\rho_k)) = \alpha\,(e_j|\rho_i) + \beta\,(e_j|\rho_k)$, where for ease of notation we suppress the system type index A. All effects $(e_j| \in \mathsf{Eff}(B)$ thus define a linear functional on the set of states $|\rho_i)_A \in \mathsf{St}(A)$, for example $(e_j|$ defines the function $\hat{e}_j$

$$\hat{e}_j : |\rho_i) \to (e_j|\rho_i) \quad \forall\ |\rho_i) \tag{2.14}$$

The following proof is from Appendix 1 of [65], following the notation in Appendix A of [15]. The set of linear combinations of these functionals, $\hat{f} = \sum_k r_k \hat{e}_k$, $r_k \in \mathbb{R}$, extend $\mathsf{Eff}(A)$ to a real vector space which we denote $V^A := \mathsf{span}\{\hat{e}_k\}$, where $\hat{f} \in V^A$. Likewise (linear combinations of) states are linear functionals $\hat{g}$ mapping $\mathsf{Eff}(A)$ to probabilities, and we can extend the set of states to a linear vector space $V_A := \mathsf{span}(\mathsf{St}(A))$ where $\hat{g} \in V_A$, $\hat{g} = \sum_{k'} r'_{k'} \rho_{k'}$. The action of $\hat{f} \in V^A$ on $\hat{g} \in V_A$ is (by construction)

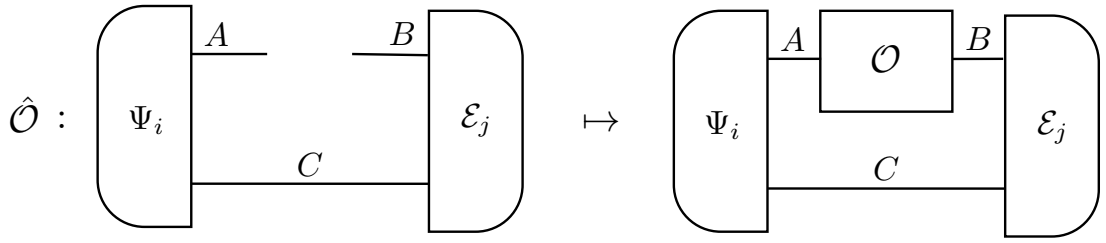
$$\hat{f}(\hat{g}) = \sum_{kk'} r_k r'_{k'} \hat{e}_k(\hat{\rho}_{k'}) = \sum_{kk'} r_k r'_{k'} \, (e_k|\rho_{k'}) \tag{2.15}$$

We now consider the more general processes with both inputs and outputs. Consider a general process with inputs A and outputs B (which can be multiple wires representing multiple systems)



(2.16)

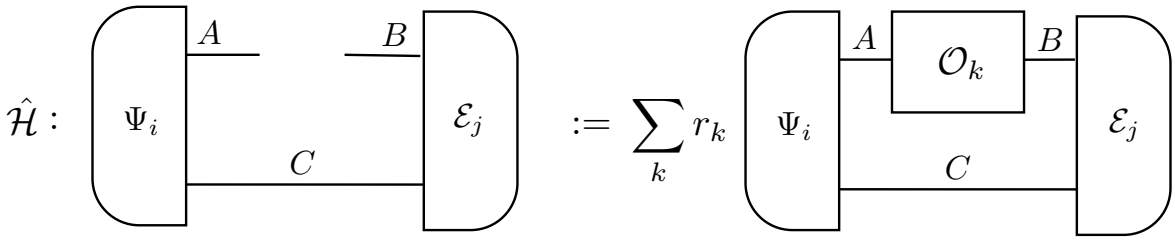
$\mathcal{O}$ is a function $\hat{\mathcal{O}}$ mapping open circuits with open wires A, B to probabilities. The most general circuit we can write down that $\mathcal{O}$ acts on is of the form



(2.17)

Any circuit involving $\mathcal{O}$ can be put in this form by noting that other processes map states to states and effects to effects and can be absorbed into Ψ_i and $\mathcal{E}_j$. Likewise, the wire C (which can represent many systems and wires) in included as $\mathcal{O}$ can be composed in parallel in a general circuit[2]. Denoting the set of processes transforming from A to B as $\mathsf{Transf}(A, B)$, we can extend this set of a real vector space as before, $V_B^A = \mathsf{span}(\hat{\mathcal{O}}_i)$ with vectors given by the functions $\hat{\mathcal{H}} = \sum_k r_k \hat{\mathcal{O}}_k$ which act on the open circuit diagrams as

[2]Later we will see this general set-up is needed to describe processes in theories that do not obey the postulate of *tomographic locality*

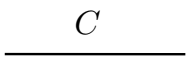


(2.18)

Thus every transformation $\mathcal{H}$ from system A to B induces a map $\hat{\mathcal{H}}_C$ from V_{AC} to V_{BC} for all systems C, as defined by

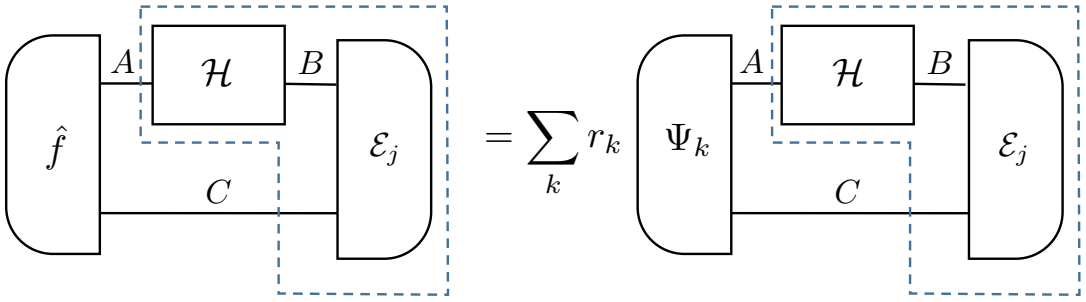
$$\hat{\mathcal{H}_C} : |\Psi)_{AC} \mapsto (\mathcal{H} \otimes \mathbb{I}) |\Psi)_{AC} \in \mathsf{St}(BC) \quad \forall\, |\Psi)_{AC} \in \mathsf{St}(AC) \tag{2.19}$$

where $\mathbb{I}$ denotes the identity transformation 'do nothing', which corresponds to the circuit element



(2.20)

We now show that, like preparations and effects being linear functionals on V_A and V^A, $\hat{\mathcal{H}}_C$ is a linear map for all systems C. Given some $\hat{f} = \sum_k r_k \hat{\Psi}_k \in V_{AB}$ the function $\hat{\mathcal{H}}_C \hat{f}$ acts on the open circuit as



(2.21)

where the equality is implied by the fact that $(\mathcal{H}\otimes\mathbb{I})\,|\mathcal{E}_j)$ is an allowed affect (depicted with the dashed line) and then applying equation 2.15. Therefore

$$\hat{\mathcal{H}}_C \hat{f} = \hat{\mathcal{H}}_C \left(\sum_k r_k \hat{\Psi}_k \right) = \sum_k r_k \mathcal{H} \otimes \mathbb{I}\,|\Psi_k) \tag{2.22}$$

which implies that $\hat{\mathcal{H}}_C$ is a linear map.

We now introduce the postulate of *finite tomography*, which is in generally assumed a priori when constructing generalized probabilistic theories, and show that it implies that the vector spaces we have constructed for our processes are finite dimensional.

Postulate 1 (Finite tomography)**.** There exists a minimal, finite set of open circuits $\{C_i\}_{i=1}^n$ s.t. for any $\hat{\mathcal{H}} \in \mathsf{Transf}(A, B)$ if

$$C_i(\mathcal{H}) = C_i \left(\sum_k r_k \mathcal{O}_k \right) \quad \forall\, i = 1, \ldots, n$$

implies $\mathcal{H} = \sum_k r_k \mathcal{O}_k$

The operational meaning of this postulate is that for any process (be it preparation, measurement of transformation), we can identify the process (e.g. perform tomography) with a finite number of experiments. For example, we can perform tomography on a preparation (state) with a finite number of measurements, or perform tomography on a process with a finite number of input states and measurements on the output. Furthermore, the postulate asserts that any two processes that cannot be differentiated from each other by finite tomography are *operationally equivalent*. This can be taken as part of the definition of an operational theory, as although the real physical details of two experimental devices can be different, they can be represented by the same mathematical object. For example, a measurement utilizing a beam splitter or an operationally equivalent experiment using a Stern-Gerlach divide, or the coherent state of an electron spin being operationally equivalent to the polarization state of a photon.

The statement of minimality implies that the set of circuits are linearly independent, $C_j \neq \sum_{k\neq j} \alpha_k C_k$ and as such the probabilities they given when composed with $\mathcal{H}$ cannot be determined from a proper subset of these probabilities (which would allow us to construct a smaller minimal set of circuits). $\{C_i\}_{i=1}^n$ is often referred to in the literature as a *fiducial* set [66].

Theorem 1. In generalized probabilistic theories obeying finite tomography V_B^A is finite dimensional

Proof. The following proof closely follows the proof outlined in section 1.2 of [40]. First we show that there exists a set of n linearly independent transformation vectors $\{\mathcal{H}_i\}_{i=1}^n$, $\mathcal{H}_i \in V_B^A$, where n is the size of the fiducial set $\{C_k\}_{k=1}^n$. To see this, first note that any vector space (finite or infinite dimensional) has a basis, and V_B^A is spanned by the set of transformations by construction, and therefore linearly independent transformations must exist. Secondly, linear independence implies that there is no set of real coefficients λ_j such that $f_k(\mathcal{H}_i) = f_k(\sum_j \lambda_j \mathcal{H}_j) = \sum_j \lambda_j f_k(\mathcal{H}_j)$ for all f_k. Therefore there are at least n such vectors, as the fiducial set $\{f_k\}_{k=1}^n$ is a minimal set.

To show that $\mathsf{dim}(V_B^A) = n$, we show that all $\mathcal{F} \in \mathsf{Transf}(A, B)$ can be written as a linear combination of these $\{\mathcal{H}_i\}_{i=1}^{n-1}$ [3]. By finite tomography, this means that for each $\mathcal{F}$ there must exist a set of real coefficients $\{\alpha_i\}_{i=1}^n$ such that

$$f_k(\mathcal{F}) = f_k\left(\sum_{i=1}^{n} \alpha_i \mathcal{H}_i\right) \quad \forall\, k \tag{2.23}$$

which can be expressed as the matrix equation

$$\begin{pmatrix} f_1(\mathcal{F}) \\ \vdots \\ f_n(\mathcal{F}) \end{pmatrix} = \begin{pmatrix} f_1(\mathcal{H}_1) & \dots & f_1(\mathcal{H}_n) \\ \vdots & \vdots & \vdots \\ f_n(\mathcal{H}_1) & \dots & f_n(\mathcal{H}_n) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \tag{2.24}$$

[3] From here we follow the clearest proof of this result that we could find, given in [40]

There is a unique solution to these equations (therefore satisfying our hypothesis) if the above matrix is invertible, which is true if the columns are linearly independent. That is, if there exists some set of λ_i such that $\sum_i \lambda_i f_k(\mathcal{H}_i) = 0 \ \forall \ k$, this linear independence would necessarily imply that $\lambda_i = 0 \ \forall \ i$. We now show that this linear independence condition follows from the linear independence of the $\{\mathcal{H}_i\}_{i=1}^n$ and finite tomography. The independence condition can be are-arranged to give

$$\begin{aligned} f_k(\mathcal{H}_1) &= -\sum_{j>1}^{n} \frac{\lambda_j}{\lambda_1} f_k(\mathcal{H}_j) \quad \forall k \\ &= f_k\left(-\sum_{j>1}^{n} \frac{\lambda_j}{\lambda_1} \mathcal{H}_j\right) \quad \forall k \end{aligned} \tag{2.25}$$

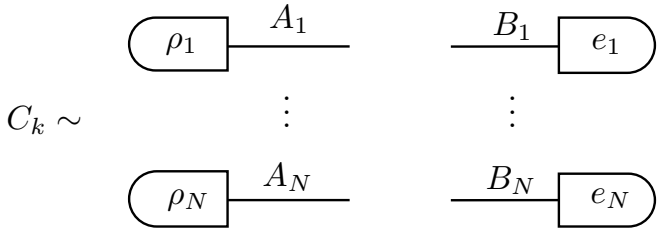
and therefore by finite tomography, $\mathcal{H}_1 = \sum_{j>1}^{n} \frac{\lambda_j}{\lambda_1} \mathcal{H}_j$. As $\mathcal{H}_i$ are linearly independent by construction, $\lambda_i = 0 \ \forall \ i$. Therefore every transformation has a unique decomposition over $\{\mathcal{H}_i\}_{i=1}^n$ which therefore form a complete basis for V_B^A, with dimension n. □

As we have seen, finite tomography constrains V_B^A to be finite dimensional, with a dimension given by the cardinality of minimal set of fiducial circuits. Consider a preparation process $|\rho)_A$, for which $V_B^A = V_A$. The dimension of V_A is therefore finite dimensional, with its dimension given by the number of *fiducial measurements* required to perform tomography on A. Likewise for a measurement process, the effect $(\mathcal{E}|_A$ belongs to vector space V^A, whose dimension is given by the size of the fiducial set of states. We can connect states in V_A with effects in V^A by the identity process (e.g. an empty wire), therefore $\mathsf{dim}(V_A) = \mathsf{dim} V^A$. V^A therefore satisfies by construction the criteria of a *dual space* of V_A, $V^A = (V_A)^*$ [60].

2.3 Tensor product composition

In quantum and classical theory, if we have two systems A and B whose local states belong to V_A and V_B respectively, any joint state of the two systems belongs to the vector space $V_A \otimes V_B$. Previously we have used $\otimes$ purely symbolically, to represent the parallel composition of two processes, but for these theories $\otimes$ is promoted to the vector tensor product. In general, we cannot assume that systems and processes will compose in this way. As we will see an addition postulate, *tomographic locality*, is required to equip a generalized probabilistic theory with this composition law. In this section we introduce tomographic locality and derive the tensor composition law. First we state the version of tomographic locality employed in the process-theory framework, which can be seen as a special case of finite tomography. It relies on the notion of a local open circuit

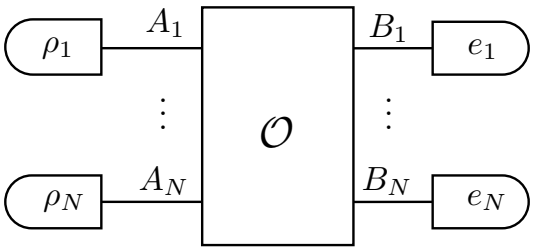
Definition 5 (Local open circuit)**.** A local open circuit is an open circuit where all inputs and outputs are connected to states and effects acting on single systems



(2.26)

Tomographic locality is the postulate that finite tomography can be performed for any process using only local process tomography. That is, we can distinguish all processes by the probabilities given by their composition with local open circuits. Operationally speaking, there exists experiments that can completely characterize a process $\mathcal{O}$ with inputs $A_1, A_2, \ldots, A_N$ and outputs $B_1, B_2, \ldots, B_M$ by varying input states on the input wires locally and performing local measurements on the output wires. Diagrammatically, the probabilities yielded by local process tomography of

$\mathcal{O}$ are of the form



(2.27)

For example, in quantum theory this would correspond to the ability to perform process tomography on a CP map without necessarily requiring the use of entangled states and effects. It is well known that quantum and classical theory obey tomographic locality [68, 15], and intuitively the violation of tomographic locality implies the existence of some non-local degrees of freedom[4]. Starting with Postulate 1, the statement of tomographic locality is as follows

Postulate 2 (Tomographic Locality)**.** For any $\mathcal{H} \in \mathsf{Transf}(A, B)$, there exists a fiducial set of local open circuits $\{L_k\}$ such that

$$L_k(\mathcal{H}) = L_k\left(\sum_i r_i \mathcal{O}_i\right) = \sum_i r_i L_k\left(\mathcal{O}_k\right)$$

implies that $\mathcal{H} = \sum_i r_i \mathcal{O}_i$

Local tomography equips a GPT with a tensor product composition given by the vector tensor product. First we illustrate this for the case of preparation processes, and show that $V_{AB} = V_A \otimes V_B$. The following proof is adapted from [40] and [15]. Consider $|\Psi) \in \mathsf{St}(AB)$, local tomography implies that the fiducial measurements are a collection of effects of the form $\{C_{ij} = |e_i)\,|e_j)\}_{(i,j)\in X\times Y}$ where $\{|e_i)\}_{i\in X}$ and $\{|e_j)\}_{j\in Y}$ are the fiducial sets of systems A and B. As $C_{ij} \in V^{AB}$ are linearly independent, there exist a set of vectors $Q_{ij} \in V_{AB}$ s.t. $(Q_{ij}|C_{kl}) = \delta_{ik}\delta_{jl}$. Furthermore

[4]This point is discussed further in the Discussion section of Chapter 4

Q_{ij} forms a basis for V_{AB} and we can choose $Q_{ij} = Q_i \otimes Q_j$ where Q_i, Q_j are similarly constructed as $(Q_i|e_k) = \delta_{ik}$ and $(Q_j|e_l) = \delta_{jl}$. As these form a basis, we can construct a linear sum of the $Q_i \otimes Q_j$ s.t.

$$\sum_{ij} r_{ij} \left(Q_i \otimes Q_j | e_k \otimes e_l\right) = (\Psi | e_k \otimes e_l) \quad \forall\, k,\, l \tag{2.28}$$

which by local tomography implies that $\Psi = \sum_{ij} r_{ij} Q_i \otimes Q_j$. As Q_i form a basis for V_A and Q_j for V_B, $\Psi \in V_A \otimes V_B \;\forall\; \Psi \in V_{AB}$ and therefore $V_{AB} = V_A \otimes V_B$. The proof is identical for effects $(\mathcal{E}|_{AB} \in V^{AB}$. For transformations $\mathcal{H} \in \mathsf{Transf}(A, B)$, the proof is identical except we construct basis Q_i for V_A and Q^j for V^B, the result being that $\mathcal{H} \in V_A \otimes V^B$. In general, a process $\mathcal{O}$ with input wires $A_1, A_2, \ldots, A_N$ and output wires $B_1, B_2, \ldots, B_M$ belongs to the vector space $V_{A_1} \otimes V_{A_2} \otimes \cdots \otimes V_{A_N} \otimes V^{B_1} \otimes V^{B_2} \otimes \cdots \otimes V^{B_M}$ (see [15, 33, 32] for detailed proofs)

2.4 Causality

Although the diagrammatic framework presented so far describes sequences of processes, composed in series and parallel, these sequences are not necessarily *causal* sequences. Diagrams naturally represent our innate perception of the directionality of time, with processes on the left of a diagram preceding those on the right. If we take this position literally, the diagrams suggest an explicit *causal ordering* to events in the experiment. This is very natural; after all the process theory framework is an operational framework for describing experiments, and experimental apparatus are typically arranged in a strict causal ordering. This is not however a prerequisite; the postulate of causality has to be put in by hand, and it is operationally possible to construct GPTs that violate our notions of causality to various degrees.

Theories with *indefinite causal ordering* have received much attention lately [8, 102, 81, 51]. One motivation for this is the belief that in quantum gravity the metric of spacetime (which dictates the ultimate causal ordering of events) will be subject

to the laws of quantum uncertainty, and hence events within that spacetime may exhibit indefinite causal ordering [66]. There are several notions of acausality, which fall broadly into the categories of *device dependent* causality (see for example [35]) and *device independent* causality (see for example [102]).

The former case refers to situations such as the use of quantum switches in quantum circuits, whereby the ordering of two unitaries depends on the state of an ancillary qubit. If the ancilla is in a superposition of the two indicator states, the unitaries are applied in a superposition of orderings. The result is a superposition of two different circuits as opposed to a single circuit with fixed orderings. In this case although processes may appear to violate a causal ordering locally, at the level the entire system there is a definite causal ordering. In the latter case, acausal correlations are correlations that violate a causal inequality [25, 7, 103, 23]. For example, causal probability distributions for two variables A and B are those that can be generated by taking a convex combination of distributions that are themselves compatible with A and B obeying a causal ordering, e.g. $A \succ B$ (A cannot signal B) or $B \succ A$ (B cannot signal A). Correlations that cannot be decomposed in this was are said to be acausal, and violate a causal inequality [7].

For this work it will be sufficient to consider an even weaker notion of causality, which is obeyed by both classical and quantum theory as well by the above cases [8][5].

Postulate 3 (Causality)**.** The probability of a preparation event is independent of the choice of measurement. E.g. for preparation process $\{\rho_k\}_{k=1}^N$, the probability weight of the preparation event ρ_k for measurement $\{e_j\}_{j=1}^M$ given by

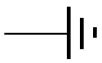
$$p_k := \sum_j (e_j|\rho_k)$$

[5]It is shown in [59] that the process matrix formalism of [102] cannot achieve signalling in time, which is possible when violating the causality axiom

is the same for all measurements on the system

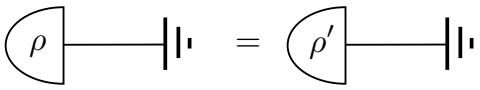
$$\sum_j (e_j|\rho_k) = \sum_i (e'_i|\rho_k) \quad \forall \{|e_j)\}, \{|e'_i)\}$$

This assumption, which for historical reasons is called the *causality postulate*, demands that the choice of a measurement has no effect on the outcome of a preparation, which can be interpreted as demanding 'no signaling from the future' [46]. This postulate can be roughly understood as following from that fact that the choice of measurement takes place in the causal future of the preparation process, and therefore cannot have any influence on the outcome statistics of the preparation[6]. It is shown in [32] that Postulate 3 is equivalent to the existence of a deterministic effect process. That is, there exists a unique process which we denote as $(u|$ or diagrammatically as



(2.29)

where the scalar given by composing this effect with a state is the same for all states, e.g. $(u|\rho) = (u|\rho')\ \forall\ \rho,\ \rho'$ or



$$\forall\ \rho,\ \rho' \in \mathsf{St}(A) \tag{2.30}$$

Lemma 1 (Characterization of causal theories (from [32]))**.** A theory is causal if and only if for every system A there exists a unique deterministic effect $(e_A|$ s.t. $(e_A|\rho_A) = 1\ \forall\ \rho_A \in \Omega_A$

Proof. Suppose there exist two deterministic effects for system A, $(e_A|$ and $(e'_A|$. Since deterministic deterministic effects have single outcomes, the causality condition (Postulate 3) gives that

[6]This also requires us to assume the postulate of Reichenbach's principle, which states that the existence of correlations observed between random variables A and B when intervening on variable A implies that there is some causal influence from A to B

$$(e_A|\rho_A) = (e'_A|\rho_A) \quad \forall\, \rho_A \in \Omega_A \tag{2.31}$$

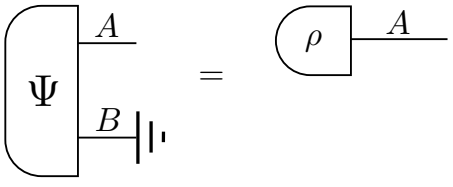
which implies that $(e_A| = (e'_A|$. Therefore all theories that obey the causality postulate satisfy the hypothesis. Next we show the converse. Suppose the deterministic effect is unique. By coarse-graining over all measurement outcomes one recovers a single-outcome effect, e.g. the unique deterministic effect, and

$$\sum_j (a_j|\rho_A) = \sum_k (b_k|\rho_A) = (e_A|\rho_A) \tag{2.32}$$

by the uniqueness of the unit effect. Therefore if a theory has unique unit effects it obeys causality □

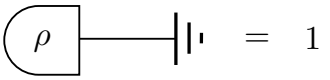
The postulate of causality imparts some useful and intuitive properties on any theory obeying it. For example, given a state belonging to two parallel-composed systems it allows us to define unique marginals on subsystems. As we can think of the above effect as a 'discarding map', this gives rise to the operational interpretation of marginal states - the states that are left over when we discard some subset of subsystems.

Definition 6 (Marginal state)**.** For $|\Psi)_{AB}$, the marginal state on system A is given by performing the discarding map on system B (given by the unique deterministic effect $\mathbb{I} \otimes |u)$).



(2.33)

The unique deterministic effect allows us to define the set of normalized states, which are all states obeying



(2.34)

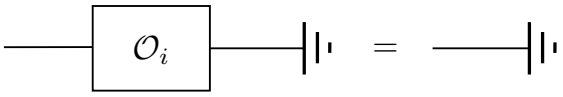
which in turn allows us to define the state space Ω_A of any given system A. Given the set of states $\mathsf{st}(A) \in V_A$ we can define the positive pointed cone generated by this set

$$K_A = \{\lambda\,|\nu)\;|\,\nu \in \mathsf{st}(A), \lambda \geq 0\} \tag{2.35}$$

and the state space is given by the interception of this cone with the unit effect

$$\Omega_A = \{|\nu)\;|\;(u|\nu) = 1,\; |\nu) \in K_A\} \tag{2.36}$$

Normalization of preparations also extends to general processes. Normalized processes obey



(2.37)

For processes with many output wires, this just amounts to applying the unique deterministic effect to each output wire[7]. Although the GPT framework allows us to consider super-normalized processes, we consider only normalized states and processes to be physically realizable. In quantum theory, this normalization condition implies that all quantum channels must be trace-preserving (or identity preserving in the Heisenberg picture). The deterministic effect in quantum theory is given by the identity operator and the inner product between states and effects is given by the trace. Likewise the normalization of effects implies that effects (given by POVM elements $\{P_i\}_{i\in X}$) sum to the identity operator $\sum_{i\in X} P_i = \mathbb{I}$. This leads to the

[7]This construction of a multi-wire discarding effect is a unit effect, and if there was a different multi-wire unit effect with would violate causality

normalization of measurements [32], which states that a measurement is physical and normalized if and only if its effects $\{(e_i|\}_{i\in X}$ obey

$$\sum_{i\in X} (e_i| = (u| \tag{2.38}$$

This is nothing more than coarse graining measurement outcomes (summing effects) and noting that, as the measurement must have some outcome, that the maximal coarse graining is a unit effect (returns probability 1). One important consequences of causality is that it implies the spatial *no-signaling principle*. This principle states that it is impossible to send information between two observers without using a signaling channel, e.g. by using only shared states and applying local processes on the subsystems that the observers have access to. Diagrammatically, no signalling in causal theories follows from the probabilities of local events factorizing over parallel wires.

Definition 7 (No-signaling principle)**.** A theory obeys the no-signaling principle if, $\forall\ |\Psi)_{AB}$ shared by Alice and Bob, and all processes $\{\mathcal{A}_i\}_{i\in X}$ locally implementable by Alice and likewise $\{\mathcal{B}_j\}_{j\in Y}$ for Bob, the probability of Alice obtaining outcome i', given by

$$p_{i'}^{(A)} = \sum_j p_{i'j}^{(AB)} \tag{2.39}$$

where

$$p_{ij}^{(AB)} = (u|_{AB}\, \mathcal{A}_i \otimes \mathcal{B}_j\, |\Psi)_{AB} \tag{2.40}$$

is the joint probability of events (i, j) for Alice and Bob, is independent of Bobs outcome j or choice of process $\{\mathcal{B}_j\}_{j\in Y}$.

The following proof is from [32]. First note that the probability of an event $\mathcal{C}_i$ acting on state $|\psi)$ is given simply by $p_i\,(u|\,\mathcal{C}_i\,|\psi)$. The probability of Alice observing event i is

$$\begin{aligned} p_i^{(A)} &= \sum_j (u|_A \, (u|_B \, \mathcal{A}_i \otimes \mathcal{B}_j \, |\Psi)_{AB} \\ &= (u|_A \left(\mathcal{A}_i \otimes \sum_j (u|_B \, \mathcal{B}_j \right) |\Psi)_{AB} \\ &= (u|_A \, \mathcal{A}_i \, |\rho)_A \end{aligned}$$

where $|\rho)_A$ is Alice's marginal state (Definition 6), and we have used the normalization of processes.

2.4.1 Recovering the convex sets formalism

In the following section we drop the 'Dirac' notation of states and effects, writing $(e| = e$ and $|\omega) = \omega$, and their inner product as $e(\omega)$ or $\omega(e)$ equivalently. The introduction of the postulates of local tomography and causality have allowed us to define the state space Ω_A for any given system A. In the convex sets framework (see for example [15]), *indeterminacy* is assumed, which means that if we can prepare two states $\rho, \rho' \in \Omega_A$ then we can prepare any convex combination of these states, e.g. $p\rho + (1-p)\rho' \in \Omega_A$, $p \in [0,1]$. For example, by flipping a coin and choosing one of two preparation processes. This extends the state space Ω_A to a convex set in V_A, and likewise K_A to a convex cone. When presenting our results we will often refer to the state space as a convex set - however our results do not rely on the assumption of indeterminacy but merely on the geometric properties of the state space Ω_A, which are equally captured by the convexified state space.

As noted in the previous section, the effects span the vector space V^A that is the dual space to the state vector space $V^A = (V_A)^*$. By construction (2.15), the effects form a cone K^A in V^A defined as

$$K^A = \{\lambda e \,|\, e \in \mathsf{Eff}(A),\, \lambda \geq 0\} \tag{2.41}$$

where $\mathsf{span}(K_A) = V_A$ and $\mathsf{span}(K^A) = V^A$. Because the pairing of a state with an effect gives a positive real number (and vice versa) by (2.15), we have the inclusions

$$K_A \subseteq (K^A)^* \tag{2.42}$$

$$K^A \subseteq (K_A)^* \tag{2.43}$$

where $(K^A)^*, (K_A)^*$ are the dual cones of K^A and K_A respectively (see [15] for examples). $(K^A)^*$ is the set of all linear functionals taking the cone of effects to the positive reals, and $(K_A)^*$ is the set of all linear functionals taking the cone of states to the positive reals. In some physical theories, the inclusion is an equality $K^A = (K_A)^*$, which means that all mathematically consistent effects are contained in the cone of effects[8]. In general this may not be the case. However the assumption that $K^A = (K_A)^*$, referred to as the *no-restriction hypothesis*, is often used as a strictly mathematical postulate [74]. Because it lacks a physical justification, we make no use of this postulate in any of our results. Violation of the no-restriction hypothesis by a theory implies that there are some effects that, although mathematically consistent, nevertheless cannot be physically realized in the theory. An example would be qubit quantum theory under the restriction that the only observables are those corresponding to elements of the Pauli group, resulting in stabilizer quantum theory [9].

The state space Ω_A is recovered by the subset of normalized states belonging to the cone K_A

$$\Omega_A = \{\omega \in K_A \,|\, u(\omega) = 1\} \tag{2.44}$$

Regardless of whether or not there exists in the theory the preparation process that

[8] by 'mathematical consistency' we refer to the effects being linear functionals that take states to the unit interval

prepares a given statistical mixture of states (by sampling a classical probability distribution to select from a set of preparation events, or preparing the mixture by taking a marginal of some entangled bipartite state), considering the convexified state space allows us to define some familiar and useful notions such as purity.

Even without assuming indeterminacy, we can define some familiar properties related to indeterminacy, such as the notion of *purity*. Purity is defined by the *refining set* of a state

Definition 8 (Refinements)**.** σ refines ρ, denoted $\sigma \succ \rho$, if there exists a set of states $\{\sigma_i\}_{i\in X}$ and a classical probability distribution $\{p_i\}_{i\in X}$ s.t.

$$\rho = \sum_i p_i \sigma_i$$

and $\sigma \in \{\sigma_i\}_{i\in X}$. The *refining set* of ρ, $\mathcal{R}(\rho)$, is the set of all σ satisfying this criteria

In indeterminate theories, $\mathcal{R}(\rho)$ has an operational interpretation, as it gives the set of states that can be used to probabilistically prepare ρ. This immediately gives us a definition of pure and mixed states. Pure states are states whose refining set contains only themselves, all other states are mixed. Refining sets form the *faces* of the convex set Ω_A [58]. The standard definition of a face from convex geometry can be phrased in terms of refining sets as follows.

Definition 9 (Faces)**.** A face F of a convex set A is a set of points $a \in A$ s.t. for any $a \in F$, if $a' \in \mathcal{R}(a)$ then $a' \in F$.

In other words, faces contain all refinements of points within the face. The set of faces of a convex set A includes the improper face (A its self), and every convex subset of A that satisfies Definition 9. The 0-dimensional faces, or vertices, trivially satisfy Definition 9. These vertices are exactly the pure states of Ω_A, as they are states that are refined only by themselves. Borrowing the language of convex sets, we will sometimes refer to them as *extremal states*. Faces can have sub-faces, and

therefore the set of faces obeys a partial ordering. This partial ordering of faces (given by the inclusion operation) is referred to in the literature as the *face lattice* of A. This is the lattice constructed by taking the set of faces $\mathsf{Faces}(A)$ under the standard set intersection $\wedge$ (the non-empty intersection of two faces is its self a face, a *sub-face* of the two intersecting faces) and the inclusion $\vee$, where $F \vee G$ denotes the *minimal face* containing faces F and G [58]. That is, $F \vee G$ is the smallest face H s.t. $F, G \subseteq H$. See Figure 2.2 for examples.

An important property of faces is whether or not they are exposed or non-exposed. An exposed face F is a face of a convex set such that there exists a linear functional $e_F(\cdot)$ on V_A, s.t.

$$e_F(\rho) = \begin{cases} \geq 0 & \forall\ \rho \in \Omega_A \\ = 0 & \forall\ \rho \in F \end{cases} \tag{2.45}$$

As linear functional equations of the form $e_F(\rho) = 0$ define hyperplanes in V_A, this can be equivalently stated as 'an exposed face has a supporting hyperplane whose intersection with the convex set is the face'. See for example figure 2.2. A non-exposed face satisfies Definition 9 but has no such supporting hyperplane. As the cone of effects K^A gives the set of linear functionals satisfying equation (2.45), the physical significance of non-exposed faces is that there cannot exist an effect that deterministically identifies the states belonging to the face. That is, there exists no effect $e_F(\cdot)$ such that $e_F(\rho) = 0$ or 1 for any $\rho \in F$. For example, as we will see in quantum and classical theory for every pure state there exists an effect that deterministically identifies it, and therefore all 0-dimensional faces are exposed, whereas in the example given in figure 2.2 there exists a pure state ρ that lies on the intersection of an edge F and a continuous curve of pure states. Therefore any effect that deterministically identifies this state also identifies all states on the edge, $e(\rho) = 0 \implies e(\omega)\ \forall\ \omega \in F$, and therefore the face given by the pure state ρ is non-exposed. Even in quantum theory there are non-exposed faces of dimension

greater than 0. For example in qutrits there are 1 dimension faces of mixed states that an non-exposed and hence cannot be measured with probability 1 [18].

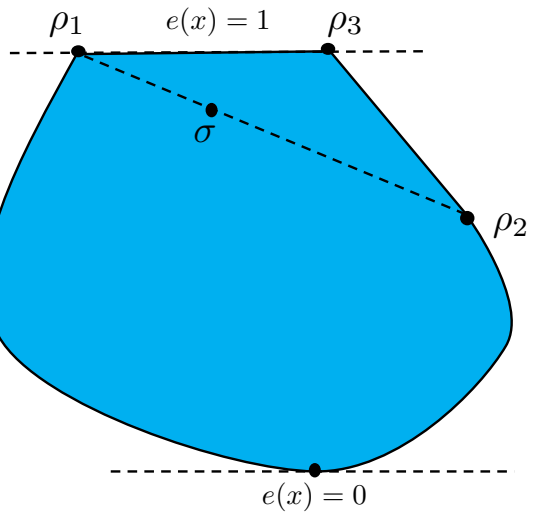


Figure 2.2: Convex set depicting a state space Ω_A. ρ_1 and ρ_2 are extremal points in the set - pure states. Note that ρ_1 is exposed, as it lies on the supporting hyperplane given by $e(x) = 1$, $x \in \Omega_A$, however ρ_2 is unexposed (there is no hyperplane that supports it, it sits between a continuous curve and an edge). The minimal face containing ρ_1 and ρ_3 is the edge $e(x) = 1$. State $\sigma = p\rho_1 + (1-p)\rho_2$, but also note that ρ_1, ρ_2, ρ_3 all refine σ (as indeed do all states in the convex set, as σ is internal).

Refinement also allows us to talk about mixedness. Mixed states are states that can be decomposed as convex combinations of other states, and therefore can (in principle) be prepared by probabilistic preparations of their refining states. The refining set of a mixed state σ gives its *minimal face*, denoted $\vee\sigma$. Given a face, the *internal states* are those which lie inside the convex set (not on the boundary), and are refined by all other states in the set. Therefore, given a face F, internal states can be defined as states that belong to F and whose minimal face is F. In general, given a state space Ω_A we we refer to the internal states of the entire convex set Ω_A as internal states, without specifying Ω_A as the face we are referring to.

The structure of quantum and classical theory allows us to define concrete notions of purity. For example, the mixedness of a qubit can be meaningfully defined by its trace-distance from the unique maximally mixed state. The existence of this state

derives from the fact that quantum and classical theory are *transitive*, in that all pure states are related by a group of transformations $\mathcal{G}$. This then induces a unique maximally mixed state μ by the Haar measure over the group of transformations

$$\mu = \int_{T \in \mathcal{G}} T(s) dT \tag{2.46}$$

for some arbitrary pure state s. However, it is known that notions of mixedness like the trace distance require a lot of addition structure to be assumed for a given generalized theory (for example, see the see the axioms required for diagonalizing states [37]). For our purposes it is sufficient to use a notion of mixedness that applies naturally to all convex operational theories without assuming additional postulates.

Definition 10 (Strict purity)**.** A state ρ is *strictly purer* than another state σ if $\rho \succ \sigma$ and $\sigma \not\succ \rho$.

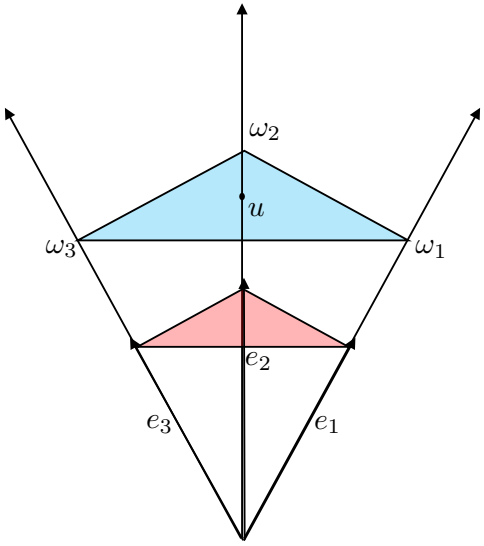
The meaning of this definition is simple: if $\rho \succ \sigma$ then σ can (at least in principle) be obtained by preparing some random preparation involving ρ. However, if $\sigma \not\succ \rho$ the converse is not true - there is no way of in principle preparing ρ using some random preparation involving σ. This is clearly the weakest and most general notion of a mixedness partial ordering, and follows naturally from the partial order given by refinements of states and therefore applies to all convex operational theories. Geometrically, ρ is strictly purer than σ if ρ belongs to a proper face of the minimal face of σ.

Having outlined the main concepts in the convex sets formalism of GPTs, we now give some simple examples.

Example 1 (Classical theory)**.** Consider a finite N dimensional classical system, where N is the number of pure states $\{\omega_j\}_{j=1\ldots N}$. These are *perfectly distinguishable*, in that there exists a single measurement with effects $\mathcal{M} = \{e_i\}_{i=1,\ldots N}$ s.t.

$$e_i(\omega_j) = \delta_{ij} \quad \forall\ i,\ j$$

As e_i are positive linear functionals on Ω_A, it follows that $\{\omega_j\}_{j=1\ldots N}$ are mutually linearly independent under all partitions, $\mathsf{span}(\mathcal{X}) \cap \mathsf{span}(\hat{\mathcal{X}}) = 0$ where $\mathcal{X}$ is some subset of the pure states and $\hat{\mathcal{X}}$ is its compliment. The convex hull of N such vectors is a simplex, with the pure states forming an orthogonal basis of V_A. Consider for example the classical trit, $N = 3$



Notice the cone of effects and effects are isomorphic $K_A \cong K^A$. As $K^A \subseteq (K_A)^*$, this implies the cones are self-dual [97]. Furthermore as the extremal states and effects correspond, the state space is *strongly self-dual*. However $\Omega_A \ncong \Omega^A$, as there are also the extremal effects $u - e_1 = e_2 + e_3$ etc, making 6 extremal effects in total, then an additional two extremal points from the effect u and the 0-effect.

The faces of an N-simplex are exposed $(N-1)$ simplexes[9]. For example the faces of the classical trit are its edges and vertices, which are classical 1 and 0 dimensional systems respectively. The pure states (vertices) are purer than the states on the edges (mixtures of two pure states) which are in turn purer than the internal states (mixtures of all three pure states) by definition 10.

Example 2 (Qutrit quantum theory)**.** Consider the state space of the quantum three-level system (qutrit). The state space is given by the set of trace 1 positive

[9]It is simple to see that the proper faces are exposed by constructing the effect that returns 0 for points only on that face. For example, the edge of the classical trit connecting the ω_0 and ω_1 pure states is picked out by the effect $u - e_2$ where e_2 picks out the state ω_2

semi-definite Hermitian operators on a three-dimensional Hilbert space - which can be represented as a convex set in an 8-dimensional real vector space [18].

The geometry of the manifold of this state space is too complicated to describe fully here. For example, unlike the qubit there are mixed states on the topological boundary, such as the state

$$\rho = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.47}$$

in the basis $\{|0\rangle\langle 0|, |1\rangle\langle 1|, |2\rangle\langle 2|\}$, which is 'picked out' by the effect $E = u - |2\rangle\langle 2|$ where u is the operator $\mathbb{I}$ and the circuit probability is given by the trace $\mathsf{tr}(E\rho)$. Furthermore there are non-exposed faces. For example, Figure 2.3 depicts a projection of the qutrit state space onto a 3-dimensional subspace [19]. The edge, drawn in red, is non-exposed as any effect that picks it out will also pick out the 'semi-circular' face below it (blue).

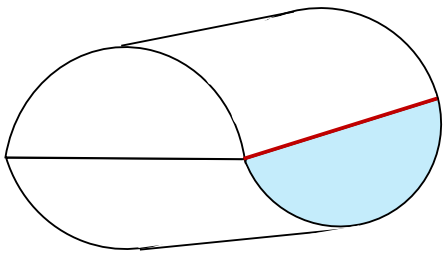


Figure 2.3: Qutrit subspace

2.4.2 Transformations

So far we have focused on preparation and measurement processes, which describe the 'kinematics' of a given operational theory. Of equal importance are the dynamics of the theory, which detail how systems evolve and interact. We have previously defined the transformation process $\mathcal{T} \in \mathsf{transf}(\mathsf{A}, \mathsf{B})$, which is a process with input

wires A and output wires B. We have shown that $\mathcal{T}$ must be a linear map from vector space V_A to vector space V_B and, as these spaces are spanned by their corresponding states Ω_A, Ω_B, $\mathcal{T}$ is uniquely defined by its action on these states (finite tomography).

$$\mathcal{T} : \Omega_A \mapsto \Omega_B; \quad \omega_a \mapsto \mathcal{T} : \omega_A \tag{2.48}$$

To achieve this, $\mathcal{T}$ must obey the following necessary conditions.

Definition 11 (Conditions for transformation operations)**.** $\mathcal{T} = \{\mathcal{T}_i\}_{i \in X}$ constitutes a physically valid map from A to B if the following conditions are satisfied for all $i \in X$

i) Positivity, $\mathcal{T}_i : \mathcal{K}_A \subseteq \mathcal{K}_B$

ii) Normalization decreasing, $u_B(T_i : \omega_A) \leq u_A(\omega_A) \; \forall \; \omega_A \in \Omega_A$

iii) Normalization preserving $\sum_k u_B(\mathcal{T}_K : \omega_A) = u_A(\omega_A) \; \forall \; \omega_A \in \Omega_A$

Condition i) is required to ensure that $\mathcal{T} : \Omega_A \mapsto \Omega_B$ is achieved with positive probability, as points that lie outside of the positive state cone yield negative probabilities on at least one effect. Condition ii) ensures that for each possible event in the map the physical states, be they normalized or sub-normalized (e.g. with an attached probability weight), are not mapped to unphysical super-normalized states. Finally, condition iii) ensures that the overall process occurs with probability 1 - one of the $i \in X$ events 'must happen'[10].

The set of physically valid maps obeying Definition 11 is in general larger than the set of physically implementable maps. For example, in reversible quantum and classical theory the dynamics are given by unitary maps and doubly stochastic maps respectively. The reversibility of these maps follows from the physical principle of dynamical time-reversibility, and imparts additional constraints on $\mathcal{T}$.

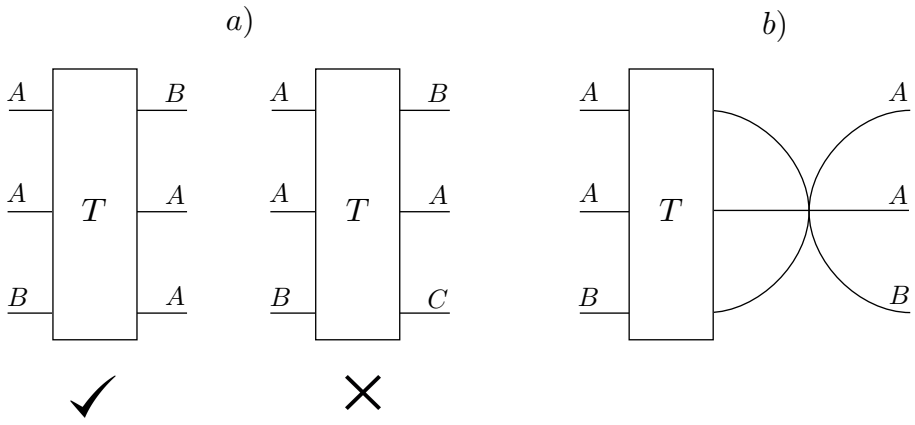
[10]note that the probability of the process 'failing' can also be encoded as an event

Definition 12 (Reversible transformations)**.** A transformation $\mathcal{T} \in \mathsf{transf}(A, B)$ is reversible iff

i) It has identical input/output wires, $B \cong A$ (up to arbitrary permutations of wires)

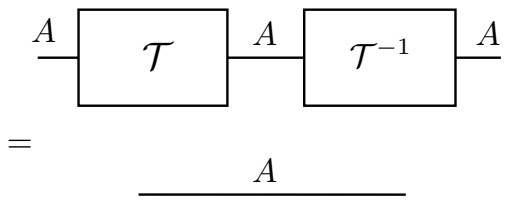
ii) $\exists$ valid $\mathcal{T}^{-1}$ s.t. $\mathcal{T}^{-1} \circ \mathcal{T} = \mathbb{I}$ where $\mathbb{I}$ is the trivial (identity) map.

The first condition states that the input and output state spaces must be isomorphic (with the isomorphism given by $\mathcal{T}$), and the second condition states that an inverse transformation exists, recovering the initial configuration of the state space from its image under $\mathcal{T}$. Graphically, these conditions can be summarized as



(2.49)

where a) shows two processes, one with mismatched outputs, and b) shows the first process being put in a form that matches inputs and outputs by permuting the output wires.

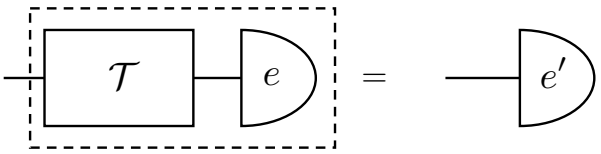


(2.50)

As the experimenter with access to transformations $\mathcal{T}_1, \mathcal{T}_2, \ldots$ is free to compose them in any order allowed by their input/outs, and by Definition 12 the reversible

transformations have compatible input/outputs, have inverses and include the identity transformation, then the set of reversible transformations forms a group $\mathcal{G}_A$. Furthermore, as the set of normalized states Ω_A is closed and compact, the action of $\mathcal{G}_A$ on Ω_A is given by a compact group. The existence of a group of reversible transformations is a very non-trivial feature, and in quantum and classical theory we will see that the existence of reversible transformations impart very strong constraints on the state spaces containing these theories.

When composed with an effect, $\mathcal{T}$ can also be seen as inducing a transformation on the effect space, e.g. $e \circ \mathcal{T} \in \Omega^A$, which follows trivially from the diagram equation



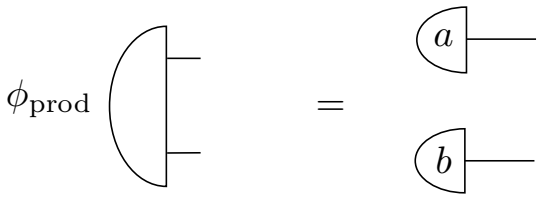
(2.51)

e.g. performing a valid transformation with an open input (state) wire and performing a measurement on all output wires is itself a valid measurement. It follows from Definition 11 and causality (Postulate 3) that $u \circ \mathcal{T} = u$, as $\mathcal{T}$ maps normalized states to normalized states and by the uniqueness of the unit effect. Furthermore, is $\mathcal{T}$ is reversible it must be a reversible transformation on all effects, $\Omega^A \circ \mathcal{T} \cong \Omega_A$.

Example 3 (Quantum theory)**.** Transformations are given by linear maps between Hilbert spaces $\Phi = \sum_k \phi_k$ where $\phi_k : \mathcal{H}_d \mapsto \mathcal{H}_{d'}$ are trace non-increasing completely positive maps. Furthermore $\sum_k \phi_k$ must be trace preserving. Likewise, these transformations can be seen as acting on the measurement operators (Heisenberg picture), with their action given by the adjoint maps $\phi_k^\dagger$. Reversible maps have the additional property of unitarity, $\Phi^\dagger = \Phi^{-1}$, and their action on $\mathcal{H}_d$ is given by the special unitary group $\mathsf{SU}(d)$.

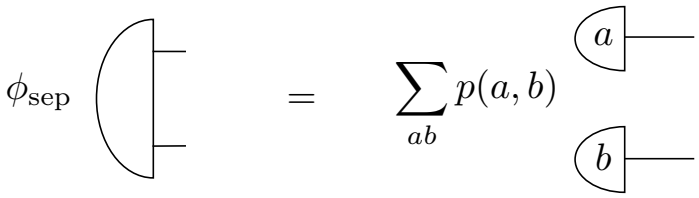
2.4.3 Composition

So far we have used basic diagrammatic reasoning coupled with the postulate of local tomography (Postulate 2), to show that the vector spaces V_A, V_B for systems A and B compose in parallel under the standard tensor product, s.t. all $\Psi \in A \otimes B$ belong to the vector space $V_A \otimes V_B$. What kinds of states populate these joint state spaces? Clearly we can construct states by sampling from the local state spaces individually or by taking correlated samples. The states sampled independently from the local distributions are called *product states*, as they factorize directly over the tensor product by construction



(2.52)

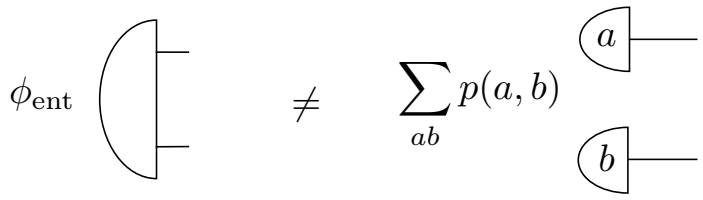
These states are the most 'local' states, as they contain no correlations between sub-systems. More generally, we can sample from the local state spaces in a dependent way. For example, if $\{a, a'\} \in \Omega_A$ and $\{b, b'\} \in \Omega_B$, we can construct a state whereby we sample from Ω_A, and then deterministically sample from Ω_B, such that $a \implies b$ and $a' \implies b'$. This results in the joint state $p\, a \otimes b + (1-p)\, a' \otimes b'$ where p is the probability of sampling a from Ω_A under our preparation procedure. These are called *separable states*, and they are convex combinations of product states.



(2.53)

Whereas product states exhibit no correlations between subsystems, separable states can carry classical correlations between sub-systems. That is, they carry the cor-

relations from the preparation sampling procedure, which we call classical as we generate them by sampling from a correlated classical probability distribution. As we know from quantum theory, separable states aren't the only states permitted by nature. Quantum systems can exhibit *entangled* states in their joint state spaces. These are states that do not belong to the set of separable states (product states belong to the set of separable states trivially).



(2.54)

Although quantum theory has these entangled states, we seek a physical justification for why nature *should* have entangled states, beyond them being mathematically possible. This is one of the questions we present an answer for in [115] and review here in Section 3.3.

The postulates of local tomography and no-signalling alone are insufficient to specify the joint state space Ω_{AB} (see for example [110]). There are multitude possible joint state spaces that satisfy local tomography. For example, we can choose to compose quantum bits such that their joint state spaces have no entangled states - this would still be a valid tensor product, but clearly it is not the only one. That said, local tomography and no-signalling are sufficient to specify the state spaces that contain the maximal and minimal number of entangled states. These extremal cases are known as the *minimal* and *maximal* tensor products.

Definition 13 (Minimal tensor product)**.** The minimal tensor product $K_{AB}^{(\min)} = K_A \otimes_{\min} K_B$ is the cone generated by the set of product states

$$K_A \otimes_{\min} K_B = \mathsf{span}\,(\omega_A \otimes \omega'_B)$$

The minimal tensor product specifies the smallest possible bipartite state space formed when combining A and B. Physically, it corresponds to a state space comprising of product states only, and given the definition of entangled state given in (2.54) represents the unique joint state space of A and B that contains no entangled states. On the other end of the spectrum, the maximal tensor product gives the largest number of entangled states [15]. This is simply the maximal set of states that can satisfy i) local tomography and ii) causality (namely, that marginal states belong to the set of allowed states on the given subsystem and iii) obeys no-signaling (which is implied by causality). As shown in [15], the maximal tensor product cone can be generated by taking the dual of the effects cones combined under the minimal tensor product

Definition 14 (Maximal tensor product)**.** $K_A \otimes_{\max} K_B = \left(K^A \otimes_{\min} K^B\right)^*$ where $K^{A,B} = (K_{A,B})^*$

In general, the joint state space Ω_{AB} is given by a set of states that lie in-between the maximal and minimal tensor products

$$K_A \otimes_{\min} K_B \subseteq K_A \otimes K_B \subseteq K_A \otimes_{\max} K_B \tag{2.55}$$

Example 4 (Classical theory)**.** It is known that classical theory is the unique GPT s.t. the maximal and minimal tensor products coincide [15]. Therefore, there is no way of combining classical systems such that they obey local tomography and causality whilst permitting entangled states.

Example 5 (Quantum theory)**.** The quantum tensor product is clearly not minimal, as bipartite quantum state spaces include entangled states. However, it is also not maximal. To see this, note that there are elements of the tensor product of two Hilbert spaces $\mathcal{H}_A \otimes \mathcal{H}_B$ that are not density matrices, but *entanglement witnesses*. These are locally positive, in that they all positive under all local POVM elements. However, there exist entangled effects for which they are negative (hence being

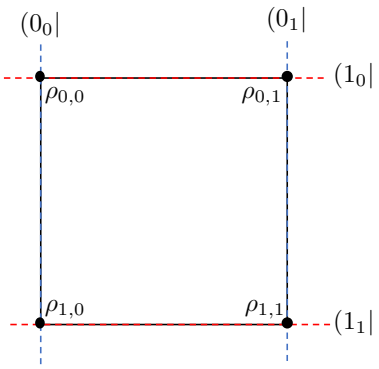
entanglement witnesses), and therefore they cannot be consistently included in the state space unless these effects are removed from the effect space [73].

Example 6 (Boxworld)**.** In [110], the question was posed 'is the no-signalling principle sufficient to constrain bipartite correlations to obey Tsirelson's bound' [38]. The authors construct a GPT known as *boxworld*, which is defined in the maximal tensor product. This GPT achieves the strongest correlations possible in the bipartite scenario with two-outcome measurements, achieving a Bell inequality violation of 4 (whereas quantum systems are constrained to have a Bell inequality violation of $2\sqrt{2}$ (Tsirelson's bound)).

In this thesis we focus on the existence of entangled states, rather than their properties and uses. For the sake of brevity, and because it is not necessary for the presentation of our results, we do not define Bell non-locality or discuss the possibility of using entangled states to violate Bell inequalities. Nor do we discuss the practical uses of entangled states, such as in quantum cryptography, super-dense coding, teleportation, quantum computing and metrology. See the review [27, 26, 54, 79] for thorough definitions and discussions of these topics. It will be sufficient for our purposes simply to operationally define entangled states as states that cannot be prepared by generating (correlated) combinations of local states.

We finish this chapter by presenting examples of how the concepts outlined in this chapter can be used to construct non-classical GPTs.

Example 7 (Polytopic theories)**.** Consider a GPT where a single system has two fiducial measurements with binary outcomes. What does this state space look like geometrically? These two fiducial measurements imply the existence of four faces of Ω_A, such that the minimal face of any two of these is Ω_A. If we assume no restriction on Ω^A, then Ω_A is identified as a four sided polytope. Denote these effect pairs $(x_a|$ where $x \in \{0,1\}$ denotes the choice of measurement and $a \in \{0,1\}$ denotes the binary measurement outcome. Ω_A has four pure states $\rho_{x,a}$ s.t. $(x'_{a'}|\rho_{x,a}) = a' \vee a$ where $\vee$ represents the Boolean OR operation. The state space is depicted below.



(2.56)

This state space is the local state space of the PR box [110], also referred to as the gbit state space [15]. Now, lets propose that these state spaces are combined under the maximal tensor product. Therefore, all joint states that give valid probabilities for local effects are included in the joint state space. Consider a joint state $|\Phi_{AB})$ such that

$$(x_a| \otimes (y_b| \, |\Phi_{AB}) = \begin{cases} 1/2 & \text{if } a \oplus b = xy \\ 0 & \text{otherwise} \end{cases} \tag{2.57}$$

It is simple to check that this state provides valid probabilities under all local measurements, and therefore such a state must belong to $\Omega_A \otimes_{\max} \Omega_B$. A version of the CHSH inequality [39], presented in [110], is

$$\sum_{a=b} p(a, b \,|\, xy = 0) + \sum_{a \neq b} p(a, b \,|\, x.y = 1) \leq 3 \tag{2.58}$$

Quantum bits can violate this inequality by achieving a LHS of $2\sqrt{2}$, whereas it is simple to check by (2.57) that the gbit achieves an LHS of 4, which is also the maximal possible by the law of total probability. Clearly $|\Phi_{AB})$ is entangled, as all separable states obey all Bell inequalities by construction. The gbit belongs to the Boxworld family [121, 110], which is the collection of GPTs that saturate the upper bound on the strength of multipartite correlations imposed by causality, local

tomography and the no-signalling principle.

What about the dynamics of this theory? In [57] it was shown that these gbits have no reversible interactions, meaning that $\exists \mathcal{T}^{-1} \implies \mathcal{T}\left(\Omega_A \otimes_{\max} \Omega_B\right) = \mathcal{T}_A\left(\Omega_A\right) \otimes_{\max} \mathcal{T}_B\left(\Omega_B\right)$. Therefore it is impossible to generate these super-quantum entangled states, or any entangled or correlated states, if we assume that all dynamics are reversible.

This has lead some to postulate that there is a trade-off between the strength of non-locality in GPTs and the richness of the reversible dynamics [57] and the possibility of protocols such as entanglement swapping, teleportation and dense coding [121]. In [4] we extended the results in [57] to all polytopic theories composed under the maximal tensor product. Consequently, even theories that are known to achieve sub-quantum violations of (2.58) (see for example [73]) are found to have no entanglement-generating reversible dynamics in the maximal tensor product. This suggests the trade-off between dynamics and entanglement is more complicated than initially hypothesized, and exploring this trade-off will be one of the topics explored in this thesis.

Chapter 3

Generalized Post-Classical Theories

In the previous chapter we outlined the framework of generalized probabilistic theories. Starting from a purely diagrammatic formalism we have recovered the convex sets formalism of state spaces, effect functionals and tensor products. In the following we will use either of these interchangeable representations of GPT's when convenient. Establishing the prior results required that we introduce certain postulates, such as local tomography and causality. We now turn our attention to a new postulate which, though strong mathematically, has a clear physical motivation. We call this postulate the *existence of a classical limit*, and to discuss it we must first discuss decoherence.

3.1 Decoherence

We postulate that in any reasonable theory of nature, systems must be able to behave classically. In some limit, the theory should be able to explain purely classical experiments. The generalized probabilistic theory framework fundamentally relies on the assumption that we have a *classical interface* with the world. We can

choose, using classical randomness and control, which experiment to perform. We can characterize states, effects and transformations in terms of the classical probability distributions that we obtain from measurements. Ultimately this classical interface should be explainable from the theory itself, rather than being an external structure we impose a priori. This is the case in quantum theory, where we can view the classical interface as an effective description of *decohered* quantum systems. Loosely speaking, quantum theory gives us classical theory 'for free' by decoherence, rather than having quantum and classical theory as two separate theories describing distinct physical systems and processes.

We postulate that any reasonable GPT should have this property too, i.e. should have an decoherence mechanism analogous to that found in quantum theory. We now consider the key features of quantum to classical decoherence which we then use to define decoherence for generalized theories.

For each quantum system $\mathcal{Q}$ with Hilbert space dimension d there is a decoherence map $D_{\mathcal{Q}}$ and classical system with state space $\Delta_{N(\mathcal{Q})}$ which is achieved under the decoherence map

$$D_{\mathcal{Q}}[\rho] := \sum_{i=1}^{d} |i\rangle\langle i|\rho|i\rangle\langle i| \tag{3.1}$$

where $\{|i\rangle\langle i|\}_{i=1,\dots d}$ is a basis of the systems Hilbert space[1].

The decoherence map describes the physical process of decoherence, which in quantum theory arises when a system interacts with an environments that is inaccessible to the experimenter. The quantum system becomes correlated with the environment in such a way that its marginal following the interaction is constrained to a sub-space that describes a classical system (simplex). There are in general many ways this can happen. To give a canonical example, consider a quantum system

[1]Note that the Hilbert space dimension is not equal to the real space dimension when represented as a GPT. The Hilbert space dimension gives the *information dimension* of the quantum state space, the largest number of perfectly distinguishable states

$|\Psi_S\rangle$ in an initial product state with an environment $|\mathcal{E}\rangle$

$$|\Psi\rangle = |\Psi_S\rangle \otimes |\mathcal{E}\rangle = \left(\sum_k \lambda_k |s_k\rangle\right) \otimes |e_0\rangle \tag{3.2}$$

where $\{|s_k\rangle\}$ is some basis for the system Hilbert space and $\{e_{k'}\}$ is some basis for the environment Hilbert space. Imagine some interaction unitary U that correlates the observables given by these bases, such that

$$U : |s_k\rangle \otimes |e_0\rangle \to |s_k\rangle \otimes |e_k\rangle \tag{3.3}$$

The post-interaction joint state is of the form

$$|\Psi'\rangle = \sum_k \lambda_k |s_k\rangle \otimes |e_k\rangle \tag{3.4}$$

and the joint density matrix of the system and environment is

$$\begin{aligned} \rho'_{SE} &= |\Psi'_{SE}\rangle\langle\Psi'_{SE}| \\ &= \sum_{kk'} \lambda_k \lambda^*_{k'} |k_S\rangle|k_E\rangle\langle k'_S|\langle k'_E| \end{aligned}$$

If we now trace out the environment (apply the unit effect), the resulting reduced state on the system is

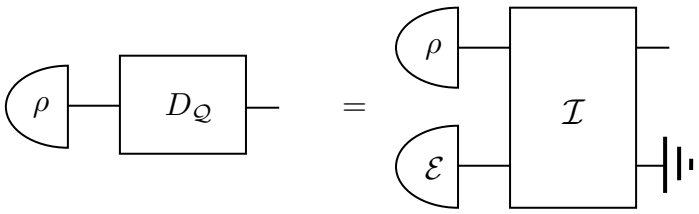
$$\rho'_S = \sum_l \langle e_l| \sum_{kk'} \lambda_k \lambda^*_{k'} |k_S\rangle|k_E\rangle\langle k'_S|\langle k'_E||e_l\rangle \tag{3.5}$$

Using the orthonormality of the environmental basis $\langle e_l|e_k\rangle = \delta_{lk}$ gives

$$\rho'_S = \sum_k |\lambda_k|^2 |s_k\rangle\langle s_k|$$

Without specifying the precise environment state or interaction unitary, we can just describe this process as a CPTP map on the system. The interaction in (3.3) is very specific to quantum theory, and we cannot expect all of its mathematical features to generalize to all theories. However, it has the following key properties which we argue should apply to any analogous decoherence process independent of the theory. Firstly, we conjecture that like quantum decoherence, generalized decoherence arises from an interaction with some environment that is then discarded.

Postulate 4 (Physical decoherence)**.** The decoherence map results from systems interacting with some inaccessible environment (e.g. an interaction followed by 'discarding' the environment system)



(3.6)

Note that we do not place any constraints on the structure of the environment or the type of interaction $\mathcal{I}$. For example, the interaction given by (3.3) results in perfect correlations between the specified system and environment basis states, but this need not be the case in general. We assume the state on the environment can be chosen to be pure[2]. The other features of general quantum decoherence maps that we adopt are detailed below.

Definition 15 (Decoherence maps)**.** Purely decoherence maps, in quantum and general theories, obey the following properties

[2]it is trivial to show that if these decoherence maps exist, we can construct decoherence maps with mixed environment states

1. Physicality: the decoherence map is a physical map, arising from a reversible interaction with some environmental system that is then discarded, and hence must satisfy all of the constraints on transformations in a GPT. In particular, it must be *linear* and map states to states.

2. Idempotence: in quantum theory the decoherence map destroys the coherences between the basis states, and so applying it a second time does nothing. Therefore, applying it twice is the same as applying it once and $D_{\mathcal{S}}[D_{\mathcal{S}}[\sigma]] = D_{\mathcal{S}}[\sigma]$ $\forall\ \sigma \in \Omega_{\mathcal{S}}$.

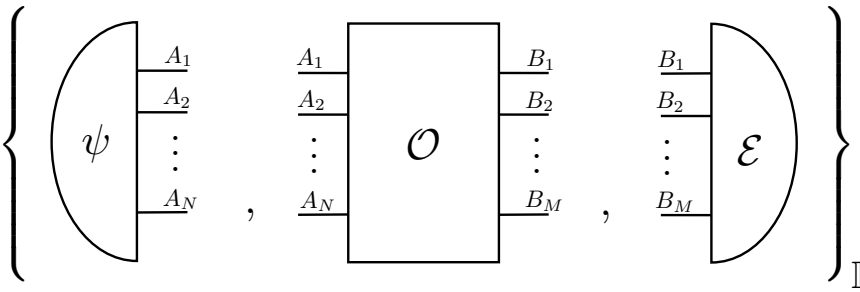
3. Purity-decreasing: the decoherence map arises from correlating with an environment and, when discarding the environment state, losing the information encoded in those correlations. As such the decoherence map cannot increase our knowledge of the input state. Therefore $D[\rho]$ cannot be strictly purer than ρ for all input states ρ. For example in quantum theory a decoherence map will not map mixed states to pure states. A state ρ is strictly purer than state σ if $\rho \succ \sigma$ and $\sigma \not\succ \rho$ and if $D_{\mathcal{S}}[\rho] \succ \rho$, then $\rho \succ D_{\mathcal{S}}[\rho]$, else $D_{\mathcal{S}}[\rho]$ is strictly purer that ρ.

It is easy to see that conditions 1, 2 and 3 are satisfied by quantum decoherence maps. In general, maps obeying these conditions could describe decoherence to any sub-theory. For example they could describe how a post-quantum theory decoheres to quantum theory [86]. This scenario is explored in Chapter 4. In this chapter, the 'sub-theory' will be classical theory.

A sub-theory is defined not only by its set of states, but also its transformations and measurements. Thus we can define the sub-theory given by a specific decoherence map as follows

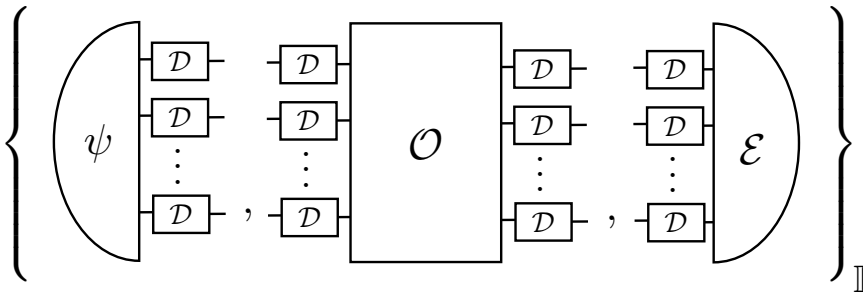
Definition 16 (Theory and sub-theory)**.** A *theory* $\mathbb{K}$ is defined by a collection of processes specifying all possible preparations, transformations and effects possible

in the theory



(3.7)

We also specify a tensor product $\otimes$ with which we can construct multipartite state spaces. Given a decoherence map $\mathcal{D}$, we can define a $\mathcal{D}$-*sub-theory* of $\mathbb{K}$ by the collection of processes given by applying the decoherence process to all wires



(3.8)

We denoted this theory $\mathbb{K}_{\mathcal{D}}$. Note that as $\mathcal{D}$ belongs to the set of allowed processes in a causal theory under Postulate 4, $\mathbb{K}_{\mathcal{D}} \subseteq \mathbb{K}$, with equality iff $\mathcal{D}$ is trivial decoherence (i.e. reversible)[3]

We are interested in decoherence maps that take systems $\mathcal{S}$ to classical systems. Returning to the quantum decoherence map, $D_{\mathcal{Q}}$ takes the quantum state space $\mathcal{Q}$ to a d-dimensional simplex (where d is the Hilbert space dimension of $\mathcal{Q}$).

[3]clearly if $\mathcal{D}$ is reversible then $\mathbb{K}_{\mathcal{D}} \cong \mathbb{K}$. We will prove triviality by showing $\mathcal{D} = \mathbb{I}$.

$$
\begin{aligned}
D_{\mathcal{Q}}[\rho] &= \sum_{i=1}^{d} |i\rangle\langle i|\rho|i\rangle\langle i| \\
&= \sum_{i=1}^{d} \langle i|\rho|i\rangle |i\rangle\langle i| \\
&= \sum_{i=1}^{d} p(i|\rho)|i\rangle\langle i|
\end{aligned}
$$

therefore all states ρ are mapped to convex mixtures over a set of perfectly distinguishable state $\{|i\rangle\langle i|\}_{i=1}^{d}$. The map is the identity map on the set of perfectly distinguishable states, and it is simple to show that the image of $\mathcal{Q}$ under $D_{\mathcal{Q}}$ is dense. Therefore the image is the convex set described by a classical simplex (see Example 1). Likewise, classical measurements can be constructed by decohering a system followed by performing a quantum measurement on it, and classical transformations are just quantum transformation with decoherence maps applies to all wires. As such, we say that *quantum theory contains a classical sub-theory*, in that all the elements of classical theory (states, transformations and measurements) can be attained by composing the processes of quantum theory with a decoherence map which its self belongs to the set of quantum processes.

Definition 17 (Classical theory)**.** (finite dimensional) Classical theory is a theory by Definition 16 whereby $\Omega_N = \Delta_N$, the effect space Ω^N obeys the no-restriction hypothesis, $\otimes = \otimes_{\min}$(necessarily, as the maximal and minimal tensor product coincide in classical theory), and the transformations on an Ω_N are given by the N permutation group

Definition 18 (Decoherence to classical theory)**.** A theory decoheres to classical theory if it has a decoherence map (Definition 5) for each system which obeys the following constraints

1. State space: the most obvious constraint is that image of the state space under

the decoherence map is a classical state space:

$$D_{\mathcal{S}}(\Omega_{\mathcal{S}}) = \Delta_{N(\mathcal{S})}$$

2. Effect space: For every effect in classical theory there is some effect in the full theory that behaves as the classical effect when we restrict to $\Delta_{N(\mathcal{S})}$. This can be formalized as for all $e_{\text{classical}}$ there exists $e \in \mathcal{E}_{\mathcal{S}}$ such that

$$e_{\text{classical}} = e \circ D_{\mathcal{S}}$$

3. (Reversible) Transformations: For any reversible classical transformation T there is a corresponding reversible post-classical transformation with the same action on the image of $D_{\mathcal{S}}$. Formally, for all classical reversible transformations $T_{\text{classical}}$ there exists some reversible $T \in \mathcal{T}_{\mathcal{S}}$ such that

$$T_{\text{classical}} = T \circ D_{\mathcal{S}}$$

4. Composites: finally, we expect decoherence to act suitably with composition, i.e. if system $\mathcal{S}_1$ decoheres to Δ_N via $D_{\mathcal{S}_1}$ and system $\mathcal{S}_2$ to Δ_M via $D_{\mathcal{S}_2}$, then the composite system $\mathcal{S}_1 \otimes \mathcal{S}_2$ can decohere to $\Delta_N \otimes \Delta_M = \Delta_{NM}$ via $D_{\mathcal{S}_1} \otimes D_{\mathcal{S}_2}$.

Example 8 (Classical sub-theory of qubit quantum theory)**.** Consider quantum theory for qubits, which decohere to classical theory for bits by applying the standard dephasing map in a given basis $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$. The decoherence map is described by

$$D_{\mathcal{S}}(\rho) := \mathsf{tr}\left(\rho|0\rangle\langle 0|\right)|0\rangle\langle 0| + \mathsf{tr}\left(\rho|1\rangle\langle 1|\right)|1\rangle\langle 1| \tag{3.9}$$

which can be achieved by performing a CNOT unitary controlled on the system, acting on an ancillary environmental qubit, followed by tracing out the environment qubit.

Measurements in the basis provide the classical measurements, with effects $|0\rangle\langle 0|$

and $|1\rangle\langle 1|$. Two qubits can decohere independently to classical bits, which can then interact as all reversible classical interactions (permutations of the join state space) can be achieved by the classical CNOT and bit flip operations, which are provided by the quantum CNOT unitary and the Pauli-X unitary. Also note that the dephasing map obeys all conditions in Definitions 15 and 18.

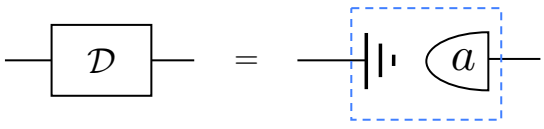
We can now define what we mean by the existence of a classical limit.

Postulate 5 (Classical limit)**.** A theory $\mathbb{K}$ has a 'classical limit' iff there exists a decoherence map $\mathcal{D}$ consistent with Definitions 15 and 18 such that we recover a classical sub-theory $\mathbb{C}_{\mathcal{D}} \subseteq \mathbb{K}$ where $\mathbb{C}_{\mathcal{D}}$ is the full classical theory satisfying Definition 17.

As is often the case when specifying a set of postulates, there are pathological cases that, whilst satisfying the criteria mathematically, do not satisfy it in spirit. For the classical limit postulate there are two such cases.

1. Trivial classical limit: $\mathcal{D}$ is the identity map (or any reversible map), resulting in $\mathbb{K} = \mathbb{C}$ (classical theory). Classical theory has a classical limit trivially.

2. Trivial decoherence. The process given by discarding a system and preparing some fixed pure state



(3.10)

   satisfies our criteria (the resulting 'classical subspace' is a 0 dimensional classical system). However, this cannot be considered true decoherence.

We call the identity map and the discarding map *trivial decoherence* and all other decoherence non-trivial. We will be interested only in theories that obtain a classical limit through non-trivial decoherence.

3.2 Decoherence in theories without entanglement

We are now in a position to explore the properties of theories obeying Postulate 5. The reason we are interested in exploring post-classical theories, i.e. theories with non-trivial classical limits, is that these are precisely the theories that can describe nature. It was known long before the advent of quantum theory that our classical understanding of the universe was incomplete, from observations such as black body radiation [109] and the stability of atomic nuclei [88]. So regardless of our knowledge or lack of knowledge of the underlying theory, we know that classical theory is only an effective description of nature that breaks down at some level. On the other hand, we know that objects *can* behave classically. The classical world that we experience must emerge somehow from some an underlying non-classical theory. One example of a theory that can explain this is quantum theory, but it is by no means the only possibility and is still arguably an abstract mathematical theory whose postulates do not have universally agreed upon physical interpretations. Could one derive quantum mechanics, or at least some partial features of quantum mechanics, simply from knowing that a classical description of the world is incomplete?

Our results mark a first step along this line of research. On the one hand, we seek a solid explanation for the counter-intuitive features of quantum theory. On the other hand, as any theory of nature should be able recover the classical limit, characterizing post-classical theories also characterizes all possible theories of nature including post-quantum theories. The ultimate goal of deriving the full formalism of quantum theory from the existence of a classical limit (and perhaps some other postulates) would be a grand result. We will focus on a more humble goal, asking the question

"*Is entanglement an inevitable feature of any post-classical theory*?"

Entanglement is perhaps one of the most counter-intuitive features of quantum theory. In the famous *Einstein Podolsk Rosen* paper [48] the authors argue against the

completeness of quantum theory, not due to the violation of local realism (discovered much later [17]), but from the very existence of entangled states. It has long been known that we can simply write down theories, other than quantum theory, that contain entangled states (e.g. [110]). However the *possibility* of entanglement is less interesting from a foundational perspective than the *necessity* for their existence. In the following we will answer the above question in the positive. Namely that

"*The existence of entangled states is necessary for the existence of a non-trivial classical limit*"

That is, without the presence of entangled states, no non-classical theory can behave classically. The existence of the classical world is contingent on the existence of entangled states - an ironic result that flies in the face of the rejections of quantum mechanics by Einstein et. al. in [48] precisely because entanglement violated their classical sensibilities. On the one hand, this just shifts the question of why entangled states exist to the question of why nature is non-classical. On the other hand, one can simply choose their favourite feature of non-classical theories (for example, the no-cloning theorem) and arrive at the same conclusion.

3.3 Deriving the existence of entangled states

We prove our result by disproving the contrapositive; that in theories without entanglement the decoherence map is a trivial decoherence map. First, we must define what we mean by theories without entanglement. Recall that the minimal tensor product (which we restate here in terms of the state cones) gives the unique method for combining locally tomographic GPT's such that there are no entangled states in the bipartite state space (Definition 13).

Min tensor product: For two systems A and B with state cones K_A and K_B respectively, the bipartite state cone formed when combining the systems under the

minimal tensor product, denoted $\boxtimes$, which is defined as

$$K_A \boxtimes K_B := \mathsf{conv}\left[\nu_A \otimes \nu_B \,|\, \nu_A \in K_A,\, \nu_B \in K_B\right] \tag{3.11}$$

e.g. it is the cone generated by the bipartite state space formed by taking all product states of the two systems. Any theory with arbitrary systems $A, B, \ldots$ that compose in parallel under the min tensor product is a theory without entanglement, and vice versa.

Another method for composing two state cones that will be important for our proof is the direct sum.

Definition 19 (Direct sum)**.** The direct sum of two cones, denoted $K_A \oplus K_B$, is given by

$$K_A \oplus K_B = \mathsf{conv}\left[\{a \oplus 0,\, 0 \oplus b \,|\, a \in K_A,\, b \in K_B\}\right]$$

where the $\oplus$ on the RHS is the standard vector direct sum and 0 is the zero-vector which gives the point of the two cones

One can visualize the direct sum as embedding the two cones in affinely independent spaces with the same $\underline{0}$ point (the point of the cone) and taking the convex hull. In quantum theory, systems composed by the direct sum represent systems with super-selection rules [137]. If for a given quantum system there is some observable that is conserved under all reversible dynamics, then in the eigenbasis of this observable all allowed density matrices obeying this conservation rule (valid preparations) must be block-diagonal, with each component within a block spanned by the eigenstates corresponding to a single value of this observable. This constraint on the density matrices arises because there cannot be coherent superpositions over these conserved quantities, and hence all off-block-diagonal amplitudes must be zero in the corresponding eigenbasis. The blocks of the density matrix generate the affinely independent state cones in the Hilbert space. We say that the state space of the quantum system *decomposes* into these disjoint subspaces, or 'super-selection sectors', where

(without loss of generality) if we allow the observable to take values $\mathcal{O} = \{0, 1, 2, \dots\}$ the statespace decomposes into $\Omega_A = \Omega_A(\mathcal{O} = 0) \oplus \Omega_A(\mathcal{O} = 1) \oplus \dots$.

These notions generalize naturally to GPT's. The spans of the super-selection sectors are affinely independent cones and the state-space cone K_A satisfies Definition 19. Conversely, in any GPT we can define a state space that decomposes as a direct sum as

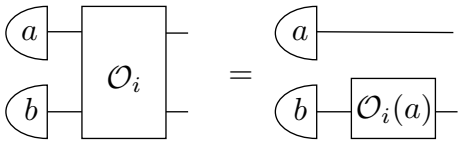
Definition 20 (Decomposable state space)**.** A state space Ω_A is decomposable iff its cone it can be written as the direct sum of its sub-cones (Def 19)

$$K_A = \bigoplus_{i=1}^{N} K_i$$

s.t. $\mathsf{span} K_i \cap \mathsf{span} K_j = 0 \ \forall \ i \neq j$

We can interpret the sub-cones that K_A decompose over as being independent subspaces, and if we label them with an index $i = 1, \dots, N$ then this index represents a 'classical degree of freedom' (a degree of freedom that cannot exhibit coherences). This gives rise to the notion of a classical observable, which is formalized nicely in the GPT framework by the notion of *non-disturbing measurements*, which are processes that act as an identity map on the system.

Definition 21 (Non-disturbing processes)**.** A process $\{\mathcal{O}_i\}_{i \in X}$ is a non-disturbing process on system A if it can be written in the form



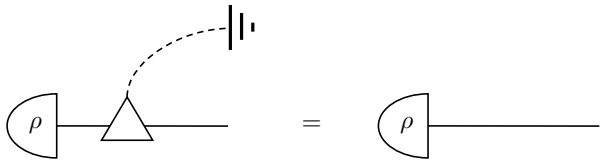
(3.12)

for all $|a) \in \Omega_A$.

Trivial non-disturbing process are where $\mathcal{O}_i(a) = \mathcal{O}_i \ \forall \ a, \ i \in X$. In this case, the environment state $|b)$ evolves independently of the system state. These processes

are not only non-disturbing, but are *non-interacting*. Examples include where $\mathcal{O}$ is the identity process, or the parallel composition of two local processes. More general non-disturbing processes can exist, and these by construction allow the environment to gain some information about the state of the system without disturbing it. This 'information leaking' is formalized by the notion of a *leak* or *partial cloner* [119], which is defined diagrammatically as

Definition 22 (Partial cloners / Leaks)**.** A leak or partial cloner is a process with multiple output wires such that discarding all but one output wire leaves the identity map on that wire



(3.13)

Intuitively, any non-trivial partial cloner / leak allows for the construction of a non-disturbing measurement. The state of the system is undisturbed by the partial cloner / leak, whist (so long as there exists some $\mathcal{O}_i(a)$ that is dependent on the input state of the system) tomography can be applied to the environment state, recovering (partial) knowledge about the state of the system[4]. A trivial partial cloner / leak is just another word for a trivial non-disturbing process[5].

Theorem 2 (Decomposability and non-disturbing measurements (from [15]))**.** For any system A with state space Ω_A, the existence of a non-trivial non-disturbing measurement $\{\mathcal{O}_i\}_{i\in X}$ implies that the state cone K_A decomposes as

$$K_A = \bigoplus_{i\in X} K_i$$

and, likewise, the decomposability of K_A implies the existence of a (mathematically

[4]such as its inclusion or exclusion in some subset of states

[5]This is just a diagrammatically more abstract object, as it makes no explicit reference to some input ancillary system

consistent) non-disturbing measurement as described[6]. .

A good example of a theory with decomposable state spaces is classical theory. In fact, classical state spaces are *completely decomposable*, in that the cone generated by a classical simplex Δ_N with pure states s_i, $i = 1, \ldots, N$, decomposes into the direct sum of the cones generated by each individual pure state. To see this, note that in the classical state space all pure states span a 1d space that is affinely independent from the space spanned by all other pure states. Thus in classical theory there exists a non-disturbing measurement that simply 'reads off' the state of the classical system, without causing any back-action. This is the most extreme case of decomposability / non-disturbing measurements, and uniquely singles out classical state spaces as the only state spaces that are totally decomposable over all pure states. In quantum systems with super-selection rules, there exist non-disturbing measurements that check which super-selection sector a given state lies in. In a more general setting, we can define GPT's that are decomposable but non-classical (see Figure 3.1)

The non-existence of non-disturbing measurements can be seen as a generalization of the no-cloning / no-broadcasting theorems of quantum theory to more general measurements which achieve partial cloning [12], where some information on an unknown preparation can be determined without disturbing it.

A result that will be important for our proof is the distributivity of the min tensor product over the direct sum

Lemma 2. The min tensor product $\boxtimes$ distributes over the direct sum $\oplus$

$$\Omega_A \boxtimes (\Omega_B \oplus \Omega_C) = (\Omega_A \boxtimes \Omega_B) \oplus (\Omega_A \boxtimes \Omega_C)$$

[6]Note that decomposability implies the mathematical existence of a non-disturbing measurement. It may be that although this measurement is consistent, it is not included in the set of allowed measurements in the theory

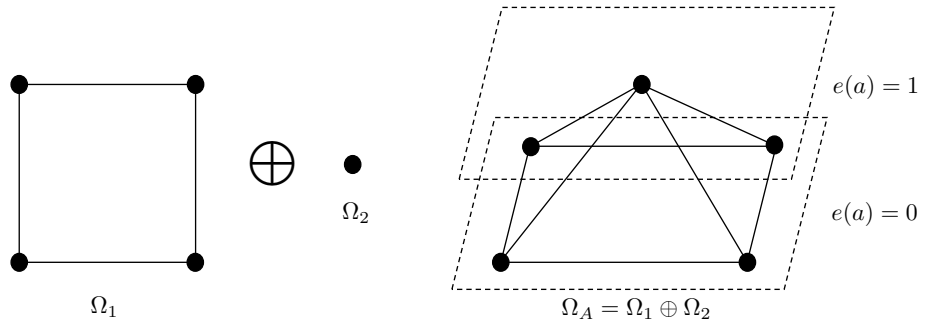


Figure 3.1: Taking the local state space of the PR-box / gbit and composing it with a 0-dimensional state space under the direct sum results in a square-based pyramid state space. Its decomposability (note that by construction $\mathsf{span}K_1 \cap \mathsf{span}K_2 = 0$) results in the existence of a 'non-disturbing measurement'. This measurement has effects e and its complementary effect $u-e$ which deterministically distinguish states on the square base with the state atthe point of the pyramid. It is possible to measure if an unkown state is in Ω_1 or Ω_2 without disturbing it. However, Ω_1 generates an indecomposable sub-cone, and thus there can be no consistent measurement that can give information on an unknown state in Ω_1 without disturbing it. As shown in [4], there are reversible interactions that correlate two such systems without altering the state of one of them, allowing non-disturbing measurements to be dynamically constructed as in Definition 21

Proof. Consider three state spaces Ω_A, Ω_B and Ω_C. We want to show that

$$\Omega_A \boxtimes (\Omega_B \oplus \Omega_C) = (\Omega_A \boxtimes \Omega_B) \oplus (\Omega_A \boxtimes \Omega_C)$$

this follows immediately by expanding both sides using Definitions 13 and 19, using the distributivity of the vector direct sum and tensor product, and noting that $0 \otimes s = 0 \otimes 0$ for all vectors s. □

Another important part of the proof concerns characterizing the reversible transformations present in post-classical theories. To do this, we must first make a few statements about the action of reversible transformations on state space. Recall that a face $f \in \mathsf{faces}(\Omega_A)$ is a convex subset of Ω_A that contains all of its refining points. First, we show that reversible transformations on the state space must map faces to isomorphic faces

Lemma 3. Reversible transformations map faces to isomorphic faces, with the isomorphism given by the reversible transformation $T(f) = f' \cong f$.

Proof. Consider some $a' \in f'$ then f' is a face if for any decomposition, $a' = pa'_1 + (1-p)a'_2$, a'_i are also in the set f'. Note that $a' = T(a)$ for some $a \in f$ and as T is reversible, this means that $a = T^{-1}(a') = T^{-1}(pa'_1 + (1-p)a'_2)$ which by linearity of T^{-1} implies that $a = pT^{-1}(a'_1) + (1-p)T^{-1}(a'_2)$. This provides a decomposition of a and as f is a face this means that $T^{-1}(a'_i) \in f$ and so a'_i are both in f'. Therefore f' is a face. It is clearly isomorphic to f as T provides the isomorphism. □

The *face lattice* [132] is the lattice formed by composing elements of the face set under the operations of intersection $\wedge$ and the join (minimal face) $\vee$. The previous result implies that the face-lattice is invariant under reversible transformations – not a surprising result, seeing as reversible transformations preserve the state space which has a fixed face lattice.

Lemma 4. A reversible transformation T on A induces an automorphism of the face lattice $\mathsf{Face}(A)$.

Proof. Lemma 3 shows that faces are mapped to faces, then reversibility of T implies that this mapping of faces must be one-to-one and therefore induces a lattice automorphism. □

An immediate corollary of this is that:

Corollary 1. $T(a_1 \vee a_2) = T(a_1) \vee T(a_2)$

That is, the minimal face of two faces is mapped to the minimal face of the two image faces. As an example, consider the local state space of the gbit, with two faces being opposing edges. The minimal face of these two edges is the entire state space, and their images under a reversible T i) are both edges (by Lemma 3) and ii) have a minimal face that is the entire state space (the square) (by Lemma 1).

We now give a simple example the illustrates the methods we will employ in our proof. In [57], it was shown that no reversible interactions exist between square

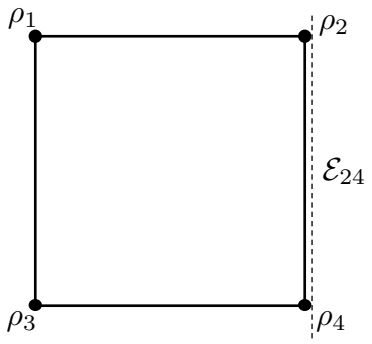


Figure 3.2: The square state space (the local state space of PR-boxes [110]) has a face lattice of extremal states ρ_i, $i = 1, \ldots, 4$, and edges which are 'picked out' by local extremal effects $\mathcal{E}_{ij}(\rho) = 0$.

state spaces, and in [5] this was generalized to arbitrary box-world state spaces. These proofs made use of a theorem from the theory of error correction [80], which states that any transformation which preserves adjacency on tuples of states can be composed using only local operations and the swapping of subsystems - and therefore cannot generate correlations between subsystems. Any such transformation is therefore non-interacting. Adjacent states are product states which differ on a single sub-systems (having a 'Hamming distance of 1'). For a transformation T that preserves adjacency, we have that for all states ρ_i, ρ_j

$$T(\rho \otimes \rho_i) = \rho_k \otimes \rho_l \implies T(\rho \otimes \rho_j) = \rho_k \otimes \rho' \quad \text{or} \quad \rho' \otimes \rho_l \tag{3.14}$$

In [57, 5] the authors note that boxworld state-spaces, combined under the maximal tensor product, have effect spaces combined under the minimal tensor product (see Section 2.4.3). As the square state space is self-dual, this describes two square effect spaces combined under the minimal tensor product. Given Corollary 1, it is simple to see that these effect spaces (and therefore the dual state spaces) cannot reversibly interact. Consider the extremal effects of the square effect space, E_1 and E_3, that do not share an edge. Consider a transformation on two adjacent extremal effects

$$T(E \otimes E_i) = E_a \otimes E_b \tag{3.15}$$

$$T(E \otimes E_j) = E_c \otimes E_d \tag{3.16}$$

Where. $E_a \neq E_c$, $E_b \neq E_d$. Consider the linear functional $\rho_a \otimes \rho_d + \rho_c \otimes \rho_d$, where ρ_x is the functional s.t. $\rho_x(E) = 0 \implies E = E_x$. This functional exists by the extremality of E (it represents an supernormalized separable state), and by its construction it returns value 0 only on the separable effects $E_a \otimes E_b$, $E_c \otimes E_d$ and their convex combinations. Therefore these image states define a face, which is an edge. As the minimal face of the pre-image effects is $E \otimes E_i \vee E \otimes E_j = E \otimes E_i \vee E_j$, which is the full effect space on subsystem 2, by Corollary 1, T cannot be reversible unless the effect space on subsystem 2 were an edge, i.e. a 2 dimensional classical system. In [4], we use this simple observation to extend the results of [57, 5] to a wide range of theories where composition is given by the minimal and maximal tensor products and the subsystems are restricted to being bit-like (all fiducial measurements have two outcomes) [4].

For more general systems than these, we require more constraining postulates than simply the existence of reversible interactions. For example, consider a subsystem that is the minimal tensor product of two convex sets, $C = A \boxtimes B$. There exists a reversible transformation $T(A' \boxtimes C) = T(A' \boxtimes A \boxtimes B) = A \boxtimes A' \boxtimes B$ where $A' \cong A$ (T is a permutation of these sub-systems A and A'). Two adjacent states in the pre-image, $a'_1 \otimes a_2 \otimes b$ and $a'_1 \otimes a_3 \otimes b$, need not be mapped to adjacent states in the image (for example if a local transformation is applied to A following the permutation). Although this may seem like cheating, a composite C is still a valid state space, and discriminating composite from non-composite systems is a difficult problem [53]. In [115] we take an different route that rules out these pathological cases. As we will show, demanding a classical limit impart strong constraints on the kind of reversible transformations you can have,

3.3.1 proof of the main result

We now present the proof of the existence of entangled states from the existence of a non-trivial decoherence map as defined in Definition 18. The following proofs derive follow those given in our publication [115]. What we prove is that any theory that can decohere to classical theory and does not have entangled states must be a composite of a classical system with a non-classical system, $\Omega_{\mathcal{S}} = \Delta_N \otimes \Omega_f$, and the decoherence map simply discards any non-classical subsystems (trivial decoherence). More succinctly,

Theories with non-trivial decoherence

must have entangled states.

First we consider the constraints decoherence places on the geometry of single systems. We show that the state space $\Omega_{\mathcal{S}}$ has the following geometric properties

3.3.2 Single-system consequences of decoherence

Result 1: *If $D_{\mathcal{S}}[\Omega_{\mathcal{S}}] = \Delta_N$ where $D_{\mathcal{S}}$ obeys Definitions 15 and 18 then the state space $\Omega_{\mathcal{S}}$ has the following properties*

1. Every classical pure state s_i is contained in a face that is its refining set f_i.
2. These faces decohere to their associated pure classical state and any state that decoheres to the classical state is in the face: $D_{\mathcal{S}}[s] = s_i \iff s \in f_i$.
3. The faces corresponding to distinct classical pure states are disjoint: $f_i \cap f_j = \emptyset \ \ \forall i \neq j$.
4. These faces are all isomorphic: $f_i \cong f_j \ \ \forall i, j$.

5. These faces are 'exposed': there exists an effect E_i such that $E_i(s) = 0 \iff s \in f_i$.

6. The minimal face containing all of these f_i is the whole state space: $\bigvee_i f_i = \mathcal{S}$

Proof. We prove these results by a series of Lemmas. Consider a state space $\Omega_{\mathcal{S}}$ with a decoherence map $D_{\mathcal{S}}$ such that $D_{\mathcal{S}}[\Omega_{\mathcal{S}}] = \Delta_N$. We label the vertices of this classical state space as $s_i \in \mathsf{Vertex}[\Delta_N]$ for $i = 1, ..., N$.

Lemma 5. Every classical pure state s_i is contained in a face f_i that is the refining set of s_i.

Proof. Follows from the definition of a face. □

Lemma 6. Every state in f_i decoheres to s_i and every state that decoheres to s_i is in f_i: $D_{\mathcal{S}}[s] = s_i \iff s_i \in f_i$.

Proof. Consider some state $s \in f_i$, this means that s is in the refining set of s_i that is, $s \succ s_i$ and $s_i = \sum_\alpha p_\alpha \sigma_\alpha$ where $\sigma_0 = s$. Then linearity and idempotence of $D_{\mathcal{S}}$, $D_{\mathcal{S}}[s_i] = \sum_\alpha p_\alpha D_{\mathcal{S}}[\sigma_\alpha]$ and therefore $D_{\mathcal{S}}[s] \succ D_{\mathcal{S}}[s_i] = s_i$. However s_i is, by definition, a vertex of the image of $D_{\mathcal{S}}$ which implies that $D_{\mathcal{S}}[s] = s_i$. Conversely, assume that $D[s] = s_i$, and for the sake of contradiction assume that $s \notin f_i$ which implies that $s \not\succ s_i = D_{\mathcal{S}}[s]$. Define a state $\tau := ps_i + (1-p)s$. Clearly $s_i \succ \tau$ but $\tau \not\succ s_i$. However $D_{\mathcal{S}}[\tau] = s_i$ and therefore $\tau \not\succ D_{\mathcal{S}}[\tau]$, $D_{\mathcal{S}}[\tau] \succ \tau$. Thus $D_{\mathcal{S}}[\tau]$ is strictly purer that τ violating the purity-decreasing postulate 3 of Definition 15. □

Lemma 7. The f_i corresponding to distinct pure classical states are disjoint: $f_i \cap f_j = \emptyset \;\; \forall\, i \neq j$.

Proof. By Lemma 6 we know that $D_{\mathcal{S}}[f_i] = s_i$ and so f_i is contained within the preimage of s_i. Hence, if the s_i and s_j are distinct then their preimages are disjoint and so $f_i \cap f_j = \emptyset$. □

Lemma 8. The f_i are all isomorphic: $f_i \cong f_j \ \ \forall\, i, j$.

Proof. In classical theory there is a reversible transformation $\pi\,[s_i] \mapsto s_j$, hence, by Postulate 3 of Definition 18 there is a reversible transformation in the post-classical theory $T_\pi\,[s_i] \mapsto s_j$. As T_π is linear it must map the refining set of s_i into the refining set of s_j and, as T_π is reversible this defines an isomorphism between f_i and f_j. □

Lemma 9. The f_i are exposed faces: there is a linear functional E_i such that $E_i(s) \geq 0\ \forall s \in \mathcal{S}$ and $E_i(s) = 0 \iff s \in f_i$.

Proof. In classical theory we have effects (i.e. linear functionals that are non-negative on the cone) E_i such that $E_i(s) \geq 0 \ \ \forall s \in \Delta_N$ and $E_i(s) = 0 \iff s = s_i$. By Postulate 2 of Definition 18 these arise from the post-classical theory as $E_i = \epsilon_i \circ D_\mathcal{S}$ and so by idempotence of decoherence we have $E_i \circ D_\mathcal{S} = E_i$. Therefore, $E_i(s) = E_i(D_\mathcal{S}[s]) \geq 0 \ \ \forall s \in \mathcal{S}$ and $E_i(s) = E_i(D_\mathcal{S}[s]) = 0 \iff D_\mathcal{S}[s] = s_i \iff s \in f_i$. □

Lemma 10. The minimal face that contains Δ_N is the entire state space: $\bigvee_{i=1}^{N} f_i = \mathcal{S}$.

Proof. Consider some state in the interior of the state space i.e. $\rho \in \mathsf{Int}[\mathcal{S}]$. This means that every other state refines ρ, i.e. $s \succ \rho \ \ \forall\ s \in \mathcal{S}$. Now consider $D_\mathcal{S}[\rho]$, and note that if a state is refined by an interior state then it too is an interior state. Therefore, for $D_\mathcal{S}[\rho]$ to not be an interior state it must be that $\rho \not\succ D_\mathcal{S}[\rho]$ but $D_\mathcal{S}[\rho] \succ \rho$, this therefore violates the purity-decreasing Postulate 3 of Definition 15. Therefore, $D_\mathcal{S}[\rho]$ must also be an interior state. As $D_\mathcal{S}[\rho] \in \Delta_N$ is interior to $\mathcal{S}$ then the minimal face that contains Δ_N must be the entire state space. □

□

3.3.3 Composite system consequences of decoherence

Next we consider the consequences of decoherence on composite systems. Essentially, as the resulting classical systems must be able to interact under classical dynamics we can use this to construct the following constraints

Result 2: Consider bipartite systems $\mathcal{S} \boxtimes \mathcal{S}$ which decohere via $D_{\mathcal{S}} \boxtimes D_{\mathcal{S}}$ to $\Delta_N \boxtimes \Delta_N = \Delta_{N^2}$. In order for the decohered systems to be able to interact via reversible classical interactions, the systems in the super-theory must obey the following constraints.

1. The classical faces in Result 1 f_i have linearly independent spans $\mathsf{Span}[f_i] \cap \mathsf{Span}[\bigcup_{k \neq i} f_k] = \{0\} \quad \forall i$
2. The minimal face containing a set of faces $\{f_i\}$ is just the convex hull of these faces, $\bigvee_i f_i = \mathsf{Conv}[\{f_i\}]$

Proof. Let us consider the bipartite state space $\mathcal{S} \boxtimes \mathcal{S}$ which decoheres via $D_{\mathcal{S}} \boxtimes D_{\mathcal{S}}$ to $\Delta_N \boxtimes \Delta_N = \Delta_{N^2}$. For example, one can consider these two state spaces decohereing independently and then being brought together.

Lemma 11. The refining set for $s_i \otimes s_j$ is $f_i \otimes f_j$.

Proof. In the minimal tensor product a generic state can be written as $\sum_\alpha p_\alpha a_\alpha \otimes b_\alpha$. Therefore, if $s_i \otimes s_j = \sum_\alpha p_\alpha a_\alpha \otimes b_\alpha$ then by considering the two marginals by applying $u \otimes \mathbb{I}$ and $\mathbb{I} \otimes u$ (i.e discarding each of the sub-systems in turn) we get that $s_i = \sum_\alpha p_\alpha a_\alpha$ and $s_j = \sum_\alpha p_\alpha b_\alpha$, and so $a_\alpha \in f_i$ and $b_\alpha \in f_j$ hence the refining set for $s_i \otimes s_j$ is just $f_i \otimes f_j$. □

Lemma 12. There exists a reversible transformation T such that

$$T :: \begin{cases} f_k \otimes f_i \mapsto f_k \otimes f_i & \forall\, i \neq j \\ f_k \otimes f_j \mapsto f_j \otimes f_j \end{cases}$$

Proof. For a classical bipartite system $\Delta_N \boxtimes \Delta_N$ there is a reversible transformation achieving the permutation of classical states

$$\pi :: \begin{cases} s_k \otimes s_i \mapsto s_k \otimes s_i & \forall\, i \neq j \\ s_k \otimes s_j \mapsto s_j \otimes s_j \end{cases}$$

therefore by Postulate 3 of Definition 18 there must be a reversible transformation T_π which has the same action as π on the image of $D_{\mathcal{S}}$. By the same argument as in the proof of Lemma 8 this must extend to the relevant refining sets as given by Lemma 11, hence taking $T = T_\pi$ we obtain our result. □

Lemma 13. The minimal face containing a set of faces $\{f_i\}_{i=1}^n$ is the convex hull of these faces, $\bigvee_{i=1}^n f_i = \mathsf{Conv}[\{f_i\}_{i=1}^n]$

Proof. We prove this result via induction, the $n = 1$ case is trivial as it simply states that $f_i = \mathsf{Conv}[\{f_i\}]$. Now for the induction, assume that $\tilde{f} := \bigvee_{i=1}^{n-1} f_i = \mathsf{Conv}[\{f_i\}_{i=1}^{n-1}]$ then the effect $\tilde{E} := \sum_{i=1}^{n-1} E_i$ is such that $\tilde{E}(s) = 0 \iff s \in \tilde{f}$. Now define the effect

$$e := E_n \otimes \tilde{E} + E_1 \otimes E_n$$

and note that $e(a \otimes b) = 0 \iff a \otimes b \in \{f_1 \otimes \tilde{f}, f_n \otimes f_n\}$. Hence, $\mathsf{Conv}[\{f_1 \otimes \tilde{f}, f_n \otimes f_n\}]$ is a face and so we have

$$f_i \otimes \tilde{f} \vee f_n \otimes f_n = \mathsf{Conv}[\{\tilde{f} \otimes f_1, f_n \otimes f_n\}].$$

Now consider the transformation T such that,

$$T :: \begin{cases} f_1 \otimes f_i \mapsto f_i \otimes f_i & i = 1, ..., n-1 \\ f_1 \otimes f_n \mapsto f_n \otimes f_n \end{cases}$$

which exists by Lem. 12. Therefore we have

$$\begin{aligned} f_1 \otimes (\tilde{f} \vee f_n) &\overset{T}{\cong} f_i \otimes \tilde{f}_i \vee f_n \otimes f_n \\ &= \mathsf{Conv}[\{f_1 \otimes \tilde{f}, f_n \otimes f_n\}] \\ &\overset{T}{\cong} f_1 \otimes \mathsf{Conv}[\{\tilde{f}, f_n\}] \end{aligned}$$

Hence we have

$$\bigvee_{i=1}^{n} f_i = \tilde{f} \vee f_n = \mathsf{Conv}[\{\tilde{f}, f_n\}] = \mathsf{Conv}[\{f_i\}_{i=1}^{n}]$$

as we required. □

Lemma 14. If the spans of f_i and f_j are equal then they must be the same face: $\mathsf{Span}[f_i] \cap \mathsf{Span}[f_j] = \mathsf{Span}[f_i] \iff i = j$.

Proof. Firstly note that as $f_i \cong f_j$ the dimension of the spans are the same, hence this is equivalent to the statement that

$$\mathsf{Span}[f_i] = \mathsf{Span}[f_j] \iff i = j.$$

The $\Leftarrow$ direction is trivial so we focus on the $\Rightarrow$ direction. Consider some state $\chi \in f_j$, by assumption we have $\chi \in \mathsf{Span}[f_i]$ and so $\chi = \sum_\alpha x_\alpha \chi_\alpha$ where $x_\alpha \in \mathbb{R}$ and $\chi_\alpha \in f_i$. Therefore $E_i(\chi) = \sum_\alpha x_\alpha E_i(\chi_\alpha) = 0$ and so $\chi \in f_i$. This is true for any state of f_i and an identical proof holds with i and j swapped, therefore $f_i = f_j$. □

Lemma 15. $\mathsf{Span}[f_i] \cap \mathsf{Span}[\bigcup_{k \neq i} f_k] = \{0\} \quad \forall i$

Proof. First note that given vector spaces U and V with subspaces $A, B \subseteq U$ and $C, D \subseteq V$ that

$$(A \otimes C) \cap (B \otimes D) = (A \cap B) \otimes (C \cap D).$$

(see for example Lemma 1.4.5 of [44]). Now consider the transformation T from

Lem. 12.

$$T :: \begin{cases} f_1 \otimes f_i \mapsto f_i \otimes f_i & 1 \neq i \\ f_1 \otimes f_k \mapsto f_1 \otimes f_k & \forall\, k \neq i \end{cases}$$

Define $A = \mathsf{Span}[f_1]$, $B := \mathsf{Span}[f_i]$ and $C := \mathsf{Span}[\bigcup_{k\neq i} f_k]$. As T is a linear reversible transformation it preserves the intersection of these subspaces. Equating the intersection of the subspaces in the pre-image and the image under T gives

$$\begin{aligned} A \otimes (B \cap C) &\overset{T}{\cong} (B \otimes B) \cap (A \otimes C) \\ &= (A \cap B) \otimes (B \cap C) \end{aligned}$$

This can only be satisfied if $A \cap B = A$ or $B \cap C = \{0\}$. The first case is ruled out by Lemma 14, whereas the second case implies that $\mathsf{Span}[f_i] \cap \mathsf{Span}[\bigcup_{k\neq i} f_k] = \{0\}$ as we require. □

□

3.3.4 Proof of main theorem

We can now combine Results 1 & 2 to prove our main theorem

Theorem 1: If a theory decoheres to classical theory (Def. 18) in the minimal tensor product, then each state space $\Omega_{\mathcal{S}}$ – which decoheres via $D_{\mathcal{S}}$ to $\Delta_{N(\mathcal{S})}$ – is of the form $\mathcal{A}_{\mathcal{S}} \boxtimes \Delta_{N(\mathcal{S})}$ and the decoherence map is $D_{\mathcal{S}} = (x_{\mathcal{S}} \circ u_{\mathcal{S}})_{\mathcal{A}_{\mathcal{S}}} \boxtimes 1_{\Delta_{N(\mathcal{S})}}$, where $x_{\mathcal{S}}$ is some fixed state of the system and $u_{\mathcal{S}}$ is the discarding map for the system.

Proof. Given the previous results we have shown that the state space can be written as $\Omega_{\mathcal{S}} = \bigvee_{i=1}^{N} f_i$ where the f_i are linearly independent. Moreover, $\bigvee_{i=1}^{n} f_i =$

$\mathsf{Conv}[\{f_i\}_{i=1}^n]$. The conjunction of these conditions implies that the cone generated by $\mathcal{S}$ decomposes into the direct sum [11]

$$\Omega_{\mathcal{S}} = \bigoplus_{i=1}^{N} f_i$$

Moreover, as the f_i are all isomorphic we can define $\mathcal{A} \cong f_i \ \ \forall\, i$. Consider the state space

$$\mathcal{S}' = \mathcal{A} \boxtimes \Delta_N$$

Using $\Delta_M = \bigoplus_{i=1}^{N} p_i$ where p_i is a zero-dimensional face, and distributivity of $\boxtimes$ over $\oplus$ (Lemma 2) we can write

$$\Omega_{\mathcal{S}'} = \bigoplus_{i=1}^{N} \mathcal{A} \boxtimes p_i$$

Finally, as p_i is zero dimensional, $p_i \boxtimes \mathcal{A} \cong \mathcal{A}$ and therefore

$$\Omega_{\mathcal{S}'} \cong \bigoplus_{i=1}^{N} \mathcal{A}_i = \Omega_{\mathcal{S}}$$

That is, $\mathcal{S}$ is just the composite of some possibly non-classical state space $\mathcal{A}$ with a classical state space Δ_N where the state of Δ_N labels which state s_i a state decoheres to. Therefore, the output state of the decoherence process depends only on the state of the classical component. Hence, can be written in the required form,

$$D_{\mathcal{S}} = (x \circ u)_{\mathcal{A}} \otimes \mathbb{I}_{\Delta_N}.$$

where u is the discarding effect and x is some fixed state of $\mathcal{A}$. □

Result 3: *For any theory that decoheres to classical theory as per Definitions 15 and 18, and whose composition rule $\otimes$ is given by the minimal tensor product $\boxtimes$, all state spaces are of the form*

$$\Omega_S = f \boxtimes \Delta_N$$

and the decoherence map is of the form

$$D_{\mathcal{S}} = (s \circ u)_f \otimes 1_{\Delta_N}$$

where u is the discarding effect and s is some fixed internal state of f. E.g. decoherence of non-classical systems comprises of discarding them.

Therefore, if we restrict ourselves to considering non-classical theories without entanglement, the only decoherence-to-classical map possible is the trivial map where we discard our non-classical systems. For example, in quantum theory this would correspond to all quantum systems regardless of their state or dimension decohereing to the zero-dimensional classical state, and the resulting classical theory being trivial.

3.3.5 Discussion of results

We have shown that if a theory has a non-trivial decoherence mechanism then the theory must have entangled states[7]. This shows that entanglement, rather than being a surprising feature of nature, is an entirely inevitable feature of any post-classical theory of nature. The results presented in the previous section are the first step towards fully characterizing all post-classical theories. For example, a natural question to ask what other features of quantum theory can be reproduced simply by demanding that the theory has a classical limit? Beyond the existence of entangled states, we expect there to be other general characteristics of this class of theories that warrant further study. There are myriad other physical features that could be implied from the existence of a classical limit such as information causality [105], bit symmetry [97] and macroscopic locality [100] to name but a few.

Another potential avenue for future work would be to see if the results presented in

[7]The following discussion, paraphrasing from [115], was originally written by myself as first author

this paper can be extended to bona fide device-independent notions of non-locality. In particular it would be interesting to determine if, perhaps by introducing additional physical postulates, it would be possible to derive the existence of non-local correlations that violate Bell inequalities. The existence of entangled states is in general a necessary but insufficient condition for observing violations of Bell inequalities. For example, non-separable states are present in the local theory of Spekken's toy model [126]. On the other hand, it has been shown that all entangled states in quantum theory display some hidden non-locality [28, 94]. By determining the additional structure present in quantum theory that gives this correspondence between entanglement and non-locality, it could be possible to derive the violation of Bell inequalities from purely physical postulates. Given the simplicity of the postulates used to derive the existence of entangled states, it is plausible that the postulates that give rise to Bell non-locality are similarly mundane.

In the next section we explore another related research line. We use the ideas laid out in the previous section to study how a quantum limit could constrain any post-quantum theory. Much like assuming a classical limit, we assume that in any deeper theory that quantum theory, systems must be able to behave like quantum systems in some limit.

Chapter 4

Hyperdecoherence

In the previous section we postulated the existence of decoherence-to-classical processes and derived the existence of entangled states. From this we can conclude that any 'post-classical' theory, that is any *extension* of classical theory to a richer theory (e.g. quantum theory) via a decoherence map, must contain entangled states. Our definition of decoherence maps (Def 15) is weak enough that it can be applied to any causal theory with reversible interactions between systems.

The proofs in the previous chapter rely on the strong constraint that the sub-theory imposes on the super-theory. The next logical step is to try and relax these constraints. Instead of making the sub-theory classical, what if we just ask that it is causal? Or that it obeys tomographic locality? This allows us to consider more general super-theory to sub-theory relations, through what we call a *hyperdecoherence* map. For example, we can define all super-theories of quantum theory as GPTs for which there exists a hyperdecoherence map (obeying Definition 15) that recover quantum theory as a sub-theory.

The exploration of possible post-quantum theories, and post-classical theories other than quantum theory, is hindered by the fact that it is very hard to construct a self-consistent theory from scratch. However, by considering this space of possible GPTs, ordered under the hyperdecoherence map, we can explore the general structure of

these theories without having to explicitly construct them. To see this, consider the space of all GPTs. This is depicted in figure **??** as a cone whose point is the trivial theory with no states, effects or transformations. The existence of a generalized decoherence map imposes a partial ordering on this space. For example, we know that quantum theory is a super-theory of classical theory. Furthermore, we know that all post-classical theories are 'entanglement theories' (i.e. theories that contain entangled states'). However, the general ordering of theories under decoherence remains unexplored. There are extensions of quantum theory, for example 'higher order interference' quantum theory [125, 122, 85, 13, 43, 14] and quartic quantum theory [138], but it is unclear if these theories can hyperdecohere to quantum theory in a physically consistent way.

In this chapter we derive a general rule that strongly constrains this partial ordering of theories. Namely that

If theory Y hyperdecoheres to theory X, $Y \succ X$, and X and Y obey causality and local tomography, then $X{=}CT$.

That is, the only sub-theory of all causal theories is classical theory. This surprising result immediately implies the following

There is no causal, pure, locally tomographic extension of quantum theory via a hyperdecoherence map.

These proofs can be seen as an extension on and improvement of the result in [86]. Before presenting our results we briefly discuss the results in [86] and the assumptions the authors use.

4.0.1 Existing no-go theorems

In [86], the authors consider all super-quantum theories that decohere to quantum theory under the same assumptions on the decoherence map as we present in

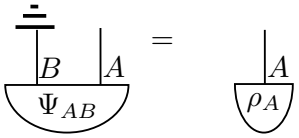
Definition 15. They also impose the following additional constraints on the set of post-quantum theories that they consider

Postulate 6 (Purity of the sub-theory)**.** All state that are pure in the sub-theory are pure in the super-theory. I.e. if we have a quantum system with state space Ω_Q and pure states $|\rho_Q) \in \Omega_Q^{\text{pure}}$, and a system in the super-theory Ω_{SQ} that decoheres to Ω_Q by $D[\Omega_{SQ}] = \Omega_Q$, then $|\rho_Q) \in \Omega_{SQ}^{\text{pure}}\ \forall\ |\rho_Q) \in \Omega_Q^{\text{pure}}$

On top of this assumption, they assume that the post-quantum theory obeys the purification postulate [32]

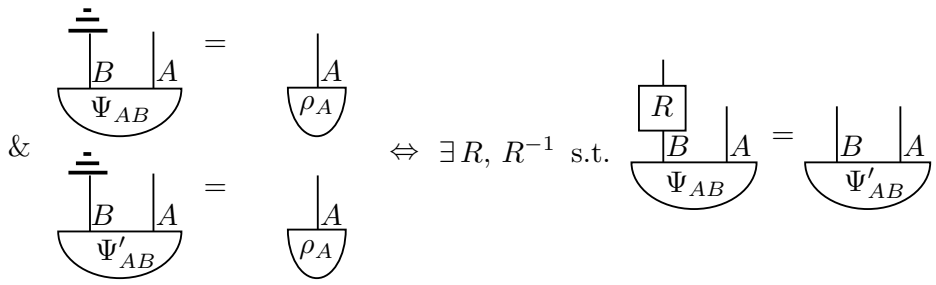
Postulate 7 (Purification)**.** A GPT obeys the purification postulate if, for every state on a single system $|\rho_A) \in \Omega_A$, there is a pure state on the bipartite composition of that system with some ancillary system B, $|\Psi_{AB}) \in \Omega_A \otimes \Omega_B$, such that

1. *purity*: $|\rho_A)$ is the marginal of $|\Psi_{AB}])$



(4.1)

2. All such 'purifications' of the state $|\rho_A)$ are connected by local reversible operations on the purifying system

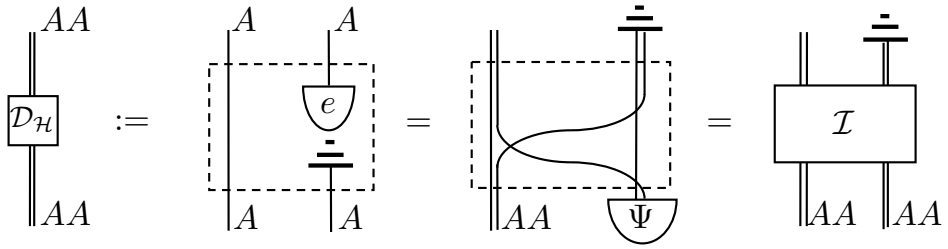


(4.2)

First, lets consider the assumption that Postulate 6 holds. This essentially asserts that pure sub-theory states (e.g. pure quantum states) are maximally informative.

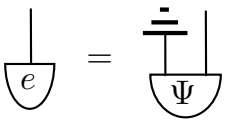
There are no states in the super-theory that refine pure quantum states. This is true for a classical sub-theory of quantum theory. For any given basis a qubit can decohere to a classical system where the pure states of the classical system correspond to the pure states of the qubit that are invariant under the decoherence map. This assumption is non-trivial, as one can easily imagine an extension of quantum theory where all quantum states are an incomplete description of the state of a system, obtained by taking a mixture of states in the super-theory.

We make use of the same assumption in our proof. The following example shows why this assumption is required - to get rid of 'fake decoherence' processes. Consider the following decoherence map, which can be constructed in quantum theory by defining the system in the super-theory be the tensor product of two identical quantum systems, $\mathcal{S}_{\text{super}} = A \otimes A$, and the system in the sub-theory be the one of the sub-systems tensored with some fixed state $|e)$, $\mathcal{S}_{\text{sub}} = A \otimes e$. The hyperdecoherence map is equivalent to simply applying the discard and prepare process to one of the sub-systems. In quantum theory, we can achieve this reversibly if the state $|e)$ is internal, as there is some entangled state that purifies it.



(4.3)

where



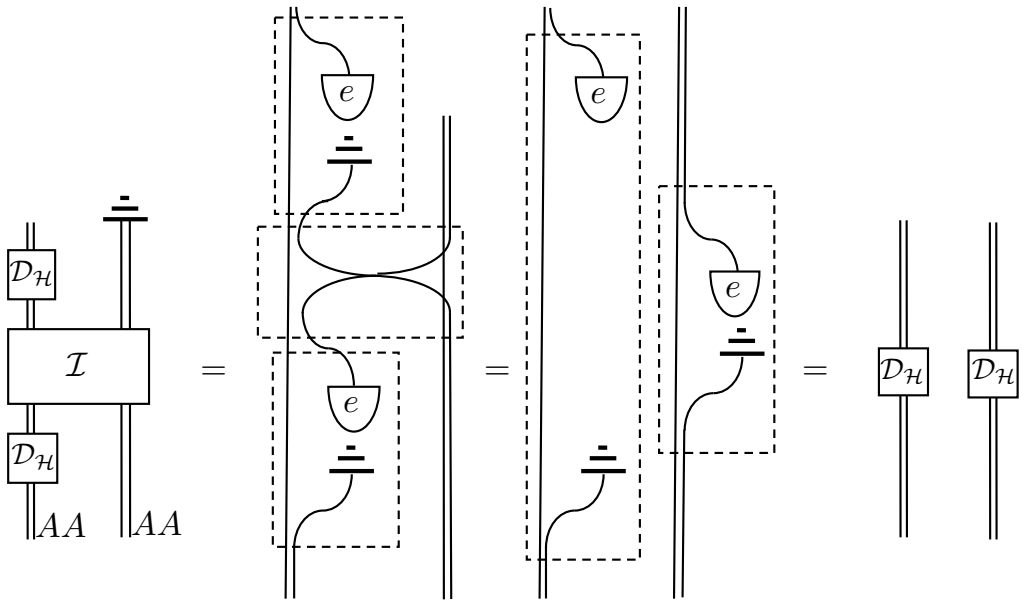
(4.4)

We achieve the hyperdecoherence map, which entails discarding the second subsystem and preparing a new state e, by performing a partial swap operation with an

environment state Ψ. As the partial swap is a reversible transformation, this is consistent with our assumptions. Furthermore it is simple to show that this map is idempotent (by composing the first term in (4.3) with its self) and causal (as quantum theory is causal). Clearly this does not constitute an interesting decoherence processes, as the sub-theory is nothing more than a state in the super-theory combined with some mixed state.

We now show that, depending on our choice of Ψ, the hyperdecoherence is trivial (there is no interacting with the environment) or the state e is mixed. In the latter case, all states in the sub theory are mixed in the super theory, and this map violates Postulate 6.

Consider the case there Ψ is a product state. It is simple to show that in this case $\mathcal{I}$ separates



(4.5)

As any convex combination of a separable process is a separable process, all non-entangled Φ result in trivial decoherence maps. In causal and locally tomographic theories, all entangled states have mixed marginals [32]. Thus, our fake hyperdecoherence satisfies all of the above assumptions except Postulate 6. It appears at first

glance that the primary function of Postulate 6 is to exclude these fake decoherence processes. We hope that in the future this postulate can be replaced by some assumption that ensures the system is not tensor-decomposable[1]. Finding postulates for the existence of sub-systems is an open problem [53], and beyond the scope of this thesis. On the other hand, there could be interesting extensions of quantum theory whereby pure quantum states are states of partial knowledge (mixed) in the super-theory.

The second assumption is quite strong, as it is shown in [33] that quantum theory can be derived from the purification postulate along with a few other assumptions. Purification is presented as the defining postulate that separates quantum theory from other possible candidates. Ideally, we would want a no-go theorem for the existence of post-quantum theories that does not assume one of the key postulates of quantum theory.

A discussion of the physical meaning of this postulate is given in [36]. The first part of the postulate states that any mixed (e.g. random) system can be seen as a sub-system of some global pure state. This is well physically motivated as it is necessary to allow for local randomness to coexist with global determinism and reversible dynamics. For example, local randomness and reversible evolution of systems are possible in classical theory, but require some initial randomness and hence a global state that is mixed. The fact that in quantum theory mixed states can be 'purified' as entangled states gets around this, allowing for global determinism alongside local randomness and reversibility. The second element of the postulate, that all purifications are related by reversible transformations on the purifying system, is a strong, mathematical assumption.

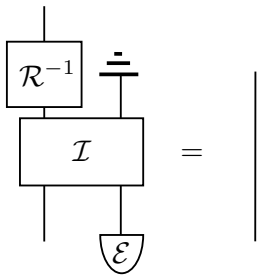
In the following section we give a proof of the same result in [86] but removing the assumption of Postulate 7. The only substantial assumptions remaining are causality, purity of the sub-theory, and local tomography.

[1]Our result will then be that the systems in sub-theories obtained through hyperdecoherence processes are tensor and/or direct sum decomposable

4.0.2 No-go theorem on hyper-decoherence

First we state our assumptions on the hyperdecoherence process. The only departure from our previous definition of decoherence in Def 15 is that we explicitly demand that the interaction process is reversible, and Postulate 6. We constrain the sub-theory to be non-classical (no non-disturbing measurements), which is true for quantum systems.

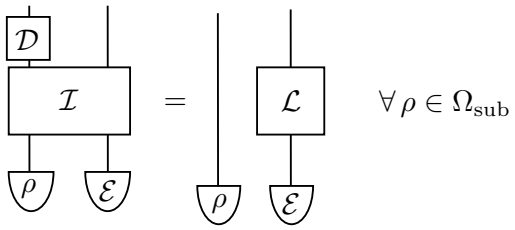
The intuition behind our proof is as follows. The decoherence map is idempotent on the states belonging to the sub-theory. Therefore it satisfies the definition of a partial cloner / leak (Definition 22). Therefore, if this interaction cannot be used to construct a non-disturbing measurement in the sub-theory (as we assume the sub-theory has no non-disturbing measurements), the partial cloner must be trivial (non-interacting). Our proof then centres around showing that, given the interaction is non-interacting for input systems belonging to the sub-theory, it must be non-interacting for input systems lying in the super-theory state space as well. Formally, we will show that, given $\mathcal{I}$ is a trivial partial cloner in Ω_{sub} it must likewise be a trivial partial cloner in the super-theory,



(4.6)

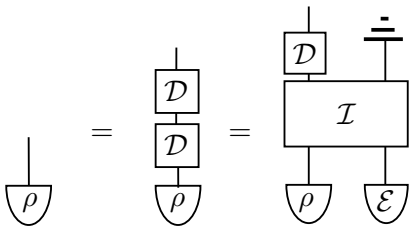
That is, the reduced transformation on the system is reversible and hence the decoherence is trivial, $\Omega_{\text{super}} \cong \Omega_{\text{sub}}$ with the isomorphism given by the the (reversible) reduced map on the system.

Lemma 16.



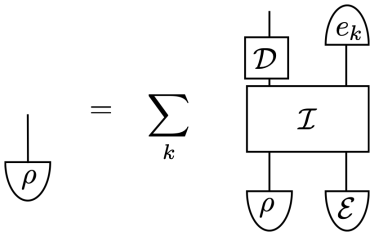
$\forall\, \rho \in \Omega_{\text{sub}}$

Proof. Starting with a pure state in the sub-theory $\rho \in \Omega_{sub}^{\text{pure}}$, idempotence implies that this state is invariant under the decoherence map. Applying the decoherence map twice gives



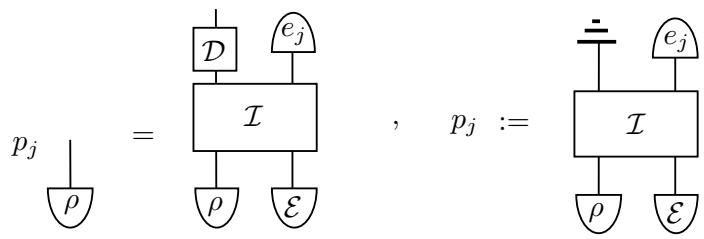
(4.7)

We can then decompose the unit effect into an arbitrary sum involving any effect $e \in \Omega^{\text{super}}$



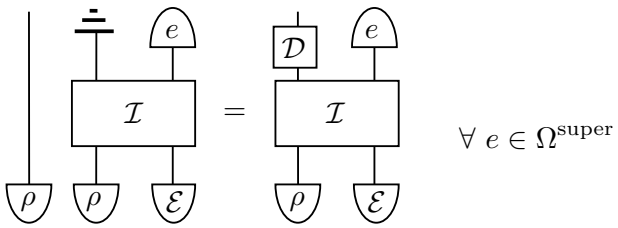
(4.8)

where



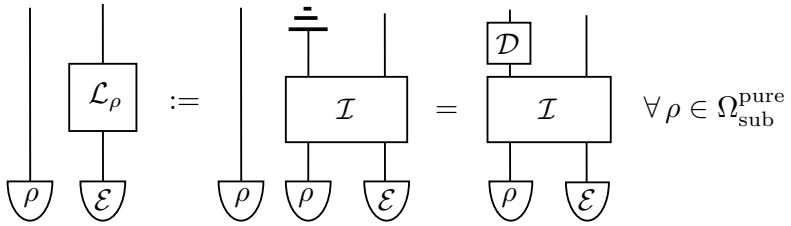
(4.9)

where we have applied the causality of the decoherence map to remove it from the definition of p_j. As ρ is pure in the sub-theory, and we have a decoherence map on the left output wire so the sub-normalized states also belong to the sub-theory, the convex sum in (4.8) must be trivial. As any $e \in \Omega^{\text{super}}$ can appear in this decomposition we have



$$\forall\ e \in \Omega^{\text{super}} \tag{4.10}$$

Applying tomographic locality then gives



$$\forall\, \rho \in \Omega^{\text{pure}}_{\text{sub}} \tag{4.11}$$

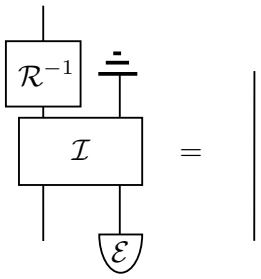
This then extends to all ρ in the sub-theory by linearity □

Lemma 17. $\mathcal{L}_\rho = \mathcal{L}\ \forall\ \rho$

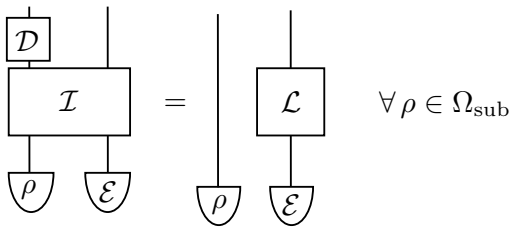
Proof. This follows immediately from the non-existence of non-trivial partial cloners on the sub-theory and the postulate of tomography. If $\mathcal{L}_\rho$ depended on the sub-theory input state then by tomography there exists some measurement that can discern and distinct images of $\mathcal{E}$ under $\mathcal{L}_\rho$. □

Result 1. If the decoherence map is a trivial partial cloner in the sub-theory, then

the decoherence map is trivial. There exists a reversible transformation $\mathcal{R}$ s.t.

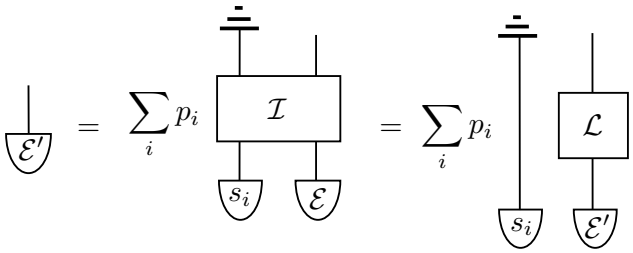


Proof. By Lemmas 16 and 17, the image of the environment state is fixed for all sub-theory inputs on the left wire



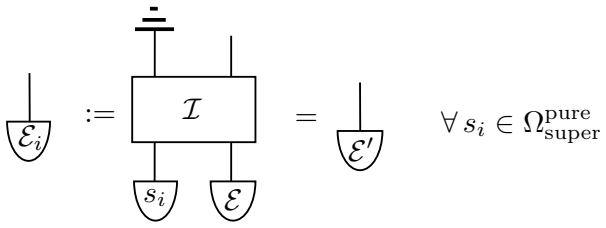
(4.12)

Let $\mathcal{L}(\mathcal{E}) = \mathcal{E}'$. First, choose $\rho \in \Omega^{\text{pure}}_{\text{super}}$ by Postulate 6. Local tomography implies that all states with pure marginals are product state (Lemma 19 of [32]). Therefore $\mathcal{I}(\rho \otimes \mathcal{E}) = \rho \otimes \mathcal{E}'$ by local tomography, then by the reversibility of $\mathcal{I}$ we have that $\mathcal{E}' \in \Omega^{\text{pure}}_{\text{super}}$. As $\mathcal{L}_\rho = \mathcal{L}$ then the image of the environment state is pure for all $\rho \in \Omega_{\text{sub}}$. By Definition 15, the sub-theory contains an internal state of the super-theory. As such we can choose $\rho \in \Omega^{\text{int}}_{\text{super}}$ and decompose the interaction as



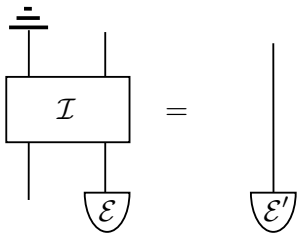
(4.13)

Where the decomposition involves arbitrary super-theory input states. As $\mathcal{E}'$ is pure, this convex sum must be trivial.



(4.14)

Therefore by local tomography we obtain



(4.15)

As the image of $\mathcal{I} :: s \otimes \epsilon$ has a pure marginal, then by local tomography it is a product state (Lemma 19 of [32]). As $\mathcal{I}$ is reversible then the image of the system under the decoherence map is isomorphic to the pre-image, e.g. $\mathcal{I}$ is a locally reversible transformation and $\Omega_{\text{super}} \cong \Omega_{\text{sub}}$

□

The consequences of Result 1 are clear. If we wish to extend a theory to a super-theory via a hyperdecoherence map, then the original theory must allow for non-trivial partial cloners, and hence non-disturbing measurements. As quantum theory, for example, does not permit non-disturbing measurements, then it cannot be extended to a theory that is causal and locally tomographic.

4.1 Discussion

Our result puts very tight constraints on what any post-quantum theory could look like, assuming it recovers quantum theory by hyperdecoherence. The necessity for

Postulate 6 are discussed in Section 4.0.1, and only plays a role of excluding fake decoherence maps. Here we briefly discuss the other two physical postulates, causality and tomographic locality, and what a super-theory that violates these postulates might look like.

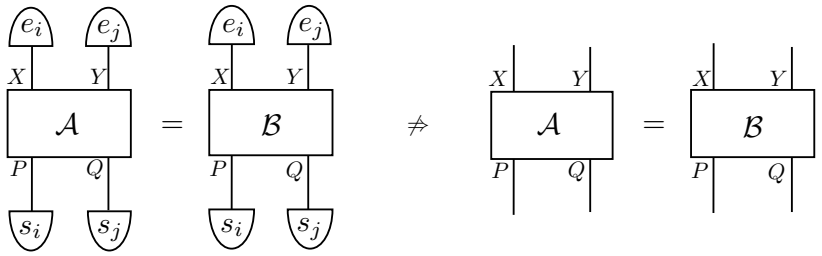
Causality: As defined in Postulate 3, causality is the assumption that preparation outcomes are independent of the choice of measurement later applied to a preparation. As we showed in Chapter 1, we can enforce causality by assuming the existence of a unique deterministic effect. What would a theory look like without causality? Acausal foil theories have been constructed; for example in [46] an acausal theory is constructed which obeys tomographic locality and determinism[2]. However, the theory violates the no-signalling principle. This is perhaps unsurprising as acausal theories can allow for 'signalling from the future', in that measurements in the causal future of a preparation can affect the outcome of that preparation. In [47] it is show that causality is required for constructing a partial ordering on events. That is, if events are time-like separated (event A precedes event B, $A \succ B$) or space-like separated ($A \not\succ B$, $B \not\succ A$). The 'past cone' of an event B is the set of all events that causally precede B, e.g. $A \succ B$. This 'no signalling from the future' principle is stated as 'the marginal probability of any event is independent of any event outside of its past cone', and it is shown that the postulate of causality is required by the no-signalling from the future postulate. However, in [59] it is shown that it is possible, at least in principle, to construct theories that violate causality and in doing so allow 'future' events to influence 'past' events, without allowing for signalling from the future to the past.

Characterizing acausal theories is an active field of research, and goes beyond the scope of this thesis. Suffice to say that causality is a fundamental physical postulate, a central assumption of the scientific method, and (as shown in [47]) is experimentally falsifiable. It remains an open question as to what reasonable acausal theories

[2]By determinism, we mean that events occur with probability 1 or 0. Thus, one can construct deterministic theories that are non-the-less acausal

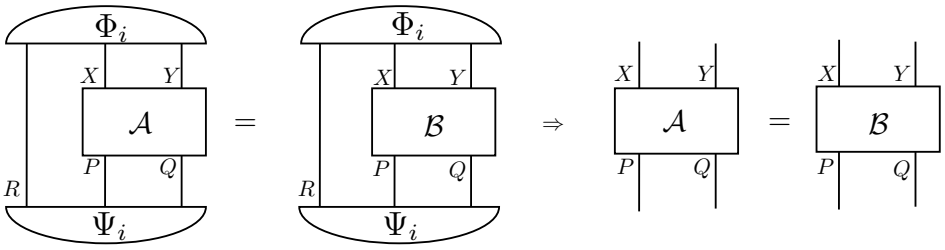
exist. Clearly if we are to relax the postulate of causality to go beyond quantum theory we can expect a significant departure from the formalism of quantum and classical theory, and the causal process theories outlined in Chapter 1.

Tomographic locality: If a theory violates tomographic locality it can still obey tomography as defined in Section 2.2. That is, performing tomography with local open circuits is insufficient to characterize a process



$\forall\, e_i \in \Omega^X,\, e_j \in \Omega^Y,\, s_i \in \Omega_P,\, s_j \in \Omega_Q$

However, if we retain the basic postulate of finite tomography then



$\forall\, R, \Psi_i \in \Omega_{RPQ}, \Phi \in \Omega^{RPQ}$

There are well known examples of GPTs that violate local tomography, probably the best known being real quantum theory [128, 78, 10, 16, 133]. Real quantum theory is constructed by replacing the complex field with the real field in standard quantum theory. As a consequence, states and effects correspond to real symmetric matrices, and transformations are completely positive maps that preserve the space of real symmetric matrices. In [14] it is shown that real quantum theory is causal and permits reversible interactions between sub-systems whilst violating tomographic locality. In [136] it is shown that real quantum theory can surpass quantum theory

in certain information transfer protocols.

An extension of quantum theory that violates local tomography would likely have some interesting properties. For example, the existence of non-local degrees of freedom. To illustrate this, we consider an example from real quantum theory (following an example given in [40]). The simplest systems in real quantum theory are *rebits*, which are the state spaces of qubits restricted to the 'equator' (i.e. the ring of real numbers). Consider some arbitrary state ρ on this ring, and the transformations $T_1(\cdot)$ and $T_2(\cdot)$ s.t.

$$T_1(\rho) = \frac{1}{2}\rho + \frac{1}{2}Y\rho Y \tag{4.16}$$

$$T_2(\rho) = \frac{1}{2}\mathbb{I} \cdot \mathrm{Tr}(\rho) \tag{4.17}$$

where Y is the Pauli-Y matrix and ρ can be written in the form

$$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \rho = \begin{pmatrix} p & q \\ q & 1-p \end{pmatrix} \tag{4.18}$$

$T_2(\cdot)$ is a 'measure-prepare' transformation that traces out the state (applies the unit effect) and prepares the state $1/2\mathbb{I}$, so clearly this is an allowed transformation. It is simple to check that $Y\rho Y$ is a real symmetric matrix trace-one for all real symmetric trace-one matrices ρ, so is an allowed transformation, and $T_1(\cdot)$ is just the mixture of this transformation with the identity map, so it an allowed transformation. It is simple to check that $T_1(\rho) = T_2(\rho)\ \forall\ \rho$ that are real symmetric and trace one, therefore these two transformations cannot be distinguished by inputting different states ρ and performing local measurements on the (identical) output states.

Consider the Bell state $|\phi^+\rangle\langle\phi^+|$ where $|\phi^\pm\rangle = 1/\sqrt{2}\,[|00\rangle \pm |11\rangle]$. Clearly this belongs to the bipartite rebit state-space by construction[3]. Consider the image of

[3]we take all states in QT with real coefficients wr.t. some basis. In this case we choose the

this state under the action of $T_1(\cdot) \otimes \mathbb{I}$ and $T_2(\cdot) \otimes \mathbb{I}$.

$$T_1 \otimes \mathbb{I}(\left|\phi^+\right\rangle \left\langle\phi^+\right|) = \frac{1}{2} \left|\phi^+\right\rangle \left\langle\phi^+\right| + \frac{1}{2} \left|\psi^-\right\rangle \left\langle\psi^-\right| \tag{4.19}$$

$$T_2 \otimes \mathbb{I}(\left|\phi^+\right\rangle \left\langle\phi^+\right|) = \frac{1}{2}\mathbb{I} \otimes \mathbb{I} \tag{4.20}$$

where $|\psi^{\pm}\rangle = 1/\sqrt{2}\,[|01\rangle \pm |10\rangle]$. Performing the two-outcome measurement with effects $\{|\phi^+\rangle\langle\phi^+| + |\psi^-\rangle\langle\psi^-|, |\phi^-\rangle\langle\phi^-| + |\psi^+\rangle\langle\psi^+|\}$ distinguishes these two states, therefore the two processes can be distinguished by non-local tomography. As noted in [40], as the image of $|\phi^+\rangle\langle\phi^+|$ are locally indiscriminable but can be non-locally discriminated, we can define a 'non-local observable'

$$Q = (T_1 \otimes \mathbb{I} - T_2 \otimes \mathbb{I})\,(\left|\phi^+\right\rangle \left\langle\phi^+\right|) \tag{4.21}$$

where, by construction, $\mathrm{Tr}\,[QE_1 \otimes E_2] = 0$ for all local measurements (real symmetric matrices E_1, E_2). One can compare this to the situation in quantum theory, where all states can be distinguished by local measurements and all observables have support over local states [68].

Of course, we cannot pretend that our treatment for extending quantum theory is totally general. There are two remaining possibilities worth mentioning. Firstly, it is quite possible that quantum theory emerges from a deeper theory via some process different to hyperdecoherence. Just because classical theory emerges from quantum theory through 'dynamical' decoherence (e.g. interactions with an environment), does not mean the same must be true for hyper-quantum theory. For example, there are extensions of quantum theory that include indefinite causal ordering [102], and it is unclear if such extensions need a dynamical decoherence mechanism to recover quantum theory, rather than recovering quantum theory under a restriction

computational basis without loss of generality.

to causal dynamics[4]. Secondly, our proofs rely on the reversibility of the interaction that gives rise to hyperdecoherence. It is possible at least in principle that there are truly irreversible processes in nature, and these could give rise to an irreversible decoherence map. However the assumption of reversibility is of fundamental importance in the theory of thermodynamics. Indeed we show in the following Chapter that without reversibility we can freely erase information, which allows us to circumvent the second law of thermodynamics.

4.2 Conclusions

In the previous chapters we have introduced the framework of generalized probabilistic theories (GPTs) - a framework general enough to encapsulate quantum theory along with any reasonable theory that can describe the outcomes of experiments. We then introduced the theory of decoherence and hyperdecoherence for generalized probabilistic theories, the first describing how GPTs can recover classical theory through dynamical decoherence (interacting with an inaccessible environment), and the second describing how we can recover more general theories (such as quantum theory) by a similar process. We then presented two new no-go theorems; the first establishing that the existence of entangled states max exist in any theory that can decohere to classical theory, and the second establishing that any theory that can hyperdecohere to quantum theory must violate either the postulate of causality, purity of the sub-theory or tomographic locality. We have then briefly discussed the implications of these no-go theories within a wider context, namely the research program of exploring possible extensions of quantum theory.

One possible avenue for future research would be to combine these two approaches. It is simple to see that, as quantum theory decoheres to classical theory, any post-quantum theory must be able to do the same. We have seen that the existence

[4]In this case, we would recover causal quantum theory because, experimentally, we are restricted to performing experiments that admit an absolute causal ordering

of a classical sub-theory imparts very strong constraints on the allowed reversible dynamics and kinematics of post-classical theories. It is likely that the same is true for the existence of a quantum sub-theory. In our second no-go theorem the only properties of quantum theory that we made use of are the non-existence of non-disturbing measurements, local tomography and causality. It is possible that stronger constraints exist if we make explicit use of the structure of quantum theory (its states, effects and transformations) as we did for classical theory in Chapter 3.

Another avenue of research would be to attempt to remove or relax the assumptions of causality and tomographic locality. Indeed the results presented in Chapter 4 currently remain unpublished due to ongoing work to remove the assumption of local tomography. The postulate of purity of the sub-theory is the first assumption we should seek to relax in future work, as it appears to only play a role in excluding some invalid hyperdecoherence processes.

Finally, the techniques developed in Chapters 3-4 could be employed to explore other non-classical features of quantum theory, and determine if these features can be derived from the existence of a classical limit. It would be fascinating if many of the features of quantum theory could be reconstructed from the existence of a classical limit alone. Or indeed if quantum theory can be singled out as the *only* post-classical theory under some assumptions such as causality and local tomography. This hypothesis is highly speculative, but we have presented strong evidence for the power of our approach. It is likely that if a full reconstruction of quantum theory is not possible by this approach, we can nevertheless recover some of its features.

In all of the results presented in the previous two chapters an indispensable ingredient has been the assumption that, fundamentally, all dynamics are reversible. In the following chapters we turn our attention to the foundations of thermodynamics in quantum and classical theory. Classical and quantum theory both obey the postulate of reversibility, and we will see that this postulate is an essential precursor to the second law of thermodynamics.

Chapter 5

Thermodynamics of individual quantum systems

5.0.1 Thermal baths

The central object that gives rise to the bounds and limitations we derive in this chapter is the thermal bath. A thermal bath is a large system, typically taken to be of infinite volume/dimension, that has reached a state of thermal equilibrium. Before we can rigorously define these statements, it is useful to give some simple descriptions of some of the basic principles describing these systems.

A microscopic state (or *microstate*) of a system is a single assignment of all of its possible degrees of freedom. For example, in a classical N-particle system, a microstate can be represented as a vector containing the instantaneous positions and momenta of all particles (and could specify additional conserved quantities such as angular momentum). Let $\mathbf{x}$ denote a microstate. As we are typically interested in counting microstates by their energy, we will often treat the energy eigenstates of the system and its microstates interchangeably. We assume that the system is closed, meaning that no energy passes in or out, and we can describe it with a single Hamiltonian. The Hamiltonian H assigns an energy to each microstate,

and we can define an energy shell as the set of states with a given energy E, $\{\mathbf{x} : H(\mathbf{x}) = E\}$. The *density of states* $\Omega(E)$ is a function giving the volume of the energy shell, $\Omega(E)dE = \mathsf{Vol}\{\mathbf{x} : E \leq H(\mathbf{x}) \leq E + dE\}$. A classic assumption in statistical mechanics is that the equilibrium state of large systems can be described by the *microcanonical ensemble*, whereby all microstate probabilities have the same probability for all microstates of the same energy[1].

$$p(\mathbf{x}) = \begin{cases} \frac{1}{\Omega(E)}, \; H(\mathbf{x}) = E \\ 0 \quad \text{otherwise} \end{cases} \tag{5.1}$$

The Shannon entropy of the microcanonical ensemble is then given by $S(E) = \ln \Omega(E)$ (we work in natural units, with the Boltzmann constant set to 1). We can trivially include an additional variable V, which represents a dimensionless volume (for example, particle number), in order to allow comparison between identical systems of different sizes, i.e. $\Omega(E, V)$. As we will see, under the assumption that a system is in thermal equilibrium with a larger system, and the global equilibrium state is described by the microcanonical distribution, we can derive that the probability of a microstate follows the Gibbs distribution [2]

Definition 23 (Gibbs distribution)**.**

$$p(\mathbf{x}) = \frac{e^{-\beta E}}{Z}, \quad Z = \sum_{\mathbf{x}} e^{-\beta E_{\mathbf{x}}} \tag{5.2}$$

[1]This is a very non-trivial assumption, that can be reached by several arguments. The most simple being to assume the assuming the *maximum entropy postulate*, whereby you assign as the equilibrium state the maximum entropy state (which, for a given energy, is the micro-canonical ensemble). Other standard arguments involve assuming ergodicity, which states that (over sufficiently large periods of time), the time spent by a system in a given region of the state space with fixed energy is proportion to the volume of that region [134]. The assumption of ergodicity allows us to equate temporal averages to ensemble averages, and the equilibrium states of the system are those that are invariant under time translations. One such invariance state is the microcanonical ensemble. However, quantum and classical systems are in general non-ergodic. This discussion goes beyond the scope of this thesis, and we take this assumption as it is a key assumption of the articles referenced in the following sections upon which this research builds.

[2]In systems that have thermalized with a bath of inverse temperature β, the distribution of microstates is typically given by the Gibbs distribution (see [62] for a resource-theoretic derivation).

We now present our definition of a thermal bath. For now we will give a very idealized definition of a thermal bath, imparting as few constraints as possible until later[3]. Note that the following definitions and proofs are taken from my publications [114] and [113].

Definition 24 (Heat bath)**.** We define a heat bath as a large system with a continuous energy spectrum defined by a Hamiltonian H_B. The density of microstates of the bath is given by $\Omega(E, V) = e^{S(E,V)}$ where $S(E, V)$ is the entropy of the microcanonical ensemble for a given energy E and (dimensionless) volume V. Furthermore, we demand that

1. The entropy is extensive, $S(kE, kV) = kS(E, V)\ \forall\ k > 0$
2. The dimensionless volume V is large
3. The bath is in a Gibbs state with inverse temperature β, s.t. a microstate with energy E has an associated probability $Z_B^{-1} e^{-\beta E}$

What is the intuition behind these assumptions? Assumptions 1 and 2 are standard assumptions for a thermal bath. The second assumption, which only becomes important later, simply states that the bath is a large system in some sense. The first assumption is non-trivial, but is certainly justified if the system is *weakly interacting*, in that the mutual information between any two partitions is negligible by comparison to the sum of the entropies of the two partitions. Then the approximation is valid by the additivity of entropy for independent distributions. The third assumption can be addressed as follows. A thermal bath is a system that has reached thermal equilibrium with an even larger system of inverse temperature β. For now, we take the limit $V \to \infty$, which we will remove in section 5.4. Consider a microstate $\mathbf{x}$ of our bath with energy $H_B(\mathbf{x})$. As the bath and the larger system

[3]This is not to say that systems not meeting our criteria cannot be considered heat baths - there are many treatments that consider thermal baths that allow for the exchange of resources other than energy [64, 61] and formalisms that allow for single quantum systems act as heat baths [124].

that contains it are together in isolation, their total energy E_0 is conserved. Therefore the larger system has energy $E_0 - H_B(\mathbf{x})$, for which there are $\Omega_S(E_0 - \mathcal{H}_B(\mathbf{x}))$ microstates that it can occupy. In other words, the number of microstates of the *combined* system, for which the thermal bath is a subsystem in microstate $\mathbf{x}$, is $\Omega_S(E_0 - H_B(\mathbf{x}))$. As the joint system is governed by the microcanonical ensemble, we can conclude (from the law of total probability) that

$$p(\mathbf{x}) = \frac{\Omega_B(E_0 - H_B(\mathbf{x}))}{\sum_{\mathbf{x}'} \Omega_B(E_0 - H_B(\mathbf{x}'))} \tag{5.3}$$

Using $\Omega(E) = e^{s(E)}$ we recover

$$\begin{aligned} p(\mathbf{x}) &\propto \Omega_B(E_0 - H_B(\mathbf{x}')) \\ &= \exp\left(S_B(E_0 - H_B(\mathbf{x}'))\right) \\ &\sim \exp\left(S_B(E_0) - \frac{\partial S_B(E)}{\partial E}\Big|_{E=E_0} E\right) \\ &\propto \exp\left(-\beta E\right) \end{aligned} \tag{5.4}$$

where we expanded the entropy as a Taylor series to first order, and used the definition of the temperature $\beta = \partial S(E)/\partial E|_{E=E_0}$. We can neglect the higher order terms in the Taylor series as (we show later) they are identical to zero in the limit $V \to \infty$. Later on we remove this 'infinite-size' approximation by including higher order terms in the Taylor series of the entropy.

5.0.2 Thermal operations

So far we have been working with a classical description of systems. The results discussed in this thesis will be essentially classical, although we will show that they port directly to semi-classical quantum systems (which we define later) and provide

upper bounds for generic quantum protocols involving coherences. In our framework, we have an (initially isolated, quantum) system with density matrix ρ_S and Hamiltonian H_S and want to determine the possible dynamics of this system when in contact with a heat bath with Hamiltonian H_B, whereby the composite system-bath system is isolated. This enforces that any interaction between system and bath, which is described by a unitary transformation U, conserves the total energy $H_S + H_B$.

Definition 25 (Energy conservation (1^{st} law))**.** The joint state of the system and bath, $\rho_S \otimes \rho_B$, evolves according to a unitary on the joint Hilbert space $H_S \otimes H_B$, which obey the commutation relation

$$[U, H_S + H_B] = 0 \tag{5.5}$$

Any such unitary can be seen as an energetically 'free' operation on the joint system-bath system, in that it requires no net injection of work into the isolated system and bath. Finding the allowed transformations of the system amounts to characterizing the set of operations that correspond to the reduced dynamics of these unitaries on the system. These operations, which we refer to as *thermal operations*, are the possible quantum maps on the system obeying Definition 25.

Definition 26 (Thermal operations on the system)**.** A quantum map $\mathcal{E}(\cdot)$ on system S in contact with a thermal bath is achievable only if

$$\mathcal{E}(\rho_S) = \mathsf{Tr}_B \left[U(\rho_S \otimes \tau_B)U^\dagger \right] \tag{5.6}$$

where U is an energy conserving unitary as per Definition 25 and τ_B is the thermal state of the bath

$$\tau_B = \frac{1}{Z} e^{-\beta H_B} \tag{5.7}$$

where H_B is he Hamiltonian of the bath, β is the inverse temperature of the bath, and Z is the partition function of the bath $Z_B = \mathsf{Tr}\left[e^{-\beta H_B}\right]$

An important corollary of our definition of thermal operations is that they leave systems that are in the thermal state of their Hamiltonian and with inverse temperature β, invariant. In other words, once a system has fully thermalized w.r.t a heat bath, there are no unitaries on the joint system (operations that require no energy import) that can change the state of the system out of its thermal state.

Corollary 2. Given a thermal operation $\mathcal{E}(\cdot)$ and a system S in thermal contact with a heat bath of inverse temperature β and in a thermal state w.r.t the baths temperature $\tau_S = \mathsf{exp}\left\{-\beta H_S\right\}$, then the state of the system is invariant under the thermal operation

$$\mathcal{E}(\tau_S) = \tau_S \tag{5.8}$$

Proof. Using the definition of the energy conserving unitary and the definition of the thermal state yields

$$\begin{aligned}[U, \tau_S \otimes \tau_B] &= \left[U, \frac{e^{-\beta H_S}}{Z_S} \otimes \frac{e^{-\beta H_S}}{Z_S}\right] \\ &= 0 \end{aligned} \tag{5.9}$$

where we have used the fact that $[U, H_S + H_B] = 0$, the diagonality of the thermal states w.r.t the joint energy eigenbasis (by definition), and that $[A, B] = 0 \leftrightarrow [A, e^B] = 0$. Therefore

$$\begin{aligned}\mathcal{E}(\tau_S) &= \mathsf{Tr}_B\left[U(\tau_S \otimes \tau_B)U^\dagger\right] \\ &= \mathsf{Tr}_B\left[(\tau_S \otimes \tau_B)\right] \\ &= 0\end{aligned}$$

□

So the thermal state of a system, with respect to the bath it is in equilibrium with, is dynamically invariant as expected. A key component of our results is that thermal operations commute with energy conserving unitaries that represent time translations of the system. This was first noted in [91], where the authors show that this compatibility imparts strong constraints on thermal operations. For example, the impossibility of creating coherent states from non-coherent states using thermal operations. For our work, it is used to show that (under the assumptions of thermal operations) we can ignore all coherences when deriving the set of thermal operations, allowing us to consider *quasi-classical states* only, which we define below.

Definition 27 (Quasi-classical states)**.** ρ_S is quasi-classical if

$$[\rho_S, H_S] = 0 \tag{5.10}$$

We now show that when characterizing thermal operations we can restrict our attention to quasi-classical states

Lemma 18 (Time translation invariance of thermal operations)**.** We say that a quantum channel on the system Hilbert space H_S represents a time translation if it is of the form

$$\mathcal{R}_t(\cdot) = e^{-iH_S t}\,(\cdot)\,e^{iH_S t} \tag{5.11}$$

Thermal operations commute with all time translations

$$\mathcal{E}\left(\mathcal{R}_t(\cdot)\right) = \mathcal{R}_t(\mathcal{E}\left(\cdot\right)) \tag{5.12}$$

Proof. We use Definition 25, and defining that at $t = 0$ (before system and bath interact) the bath is in a stationary state with respect to its Hamiltonian H_B. Composing the thermal operation with a system undergoing time translation yields

$$\begin{aligned}
\mathcal{E}\left(\mathcal{R}_t(\rho_S)\right) &= \mathsf{Tr}_B\left[Ue^{-iH_St}\rho_S e^{iH_St} \otimes \tau_B U^\dagger\right] \\
&= \mathsf{Tr}_B\left[Ue^{-iH_St}\rho_S e^{iH_St} \otimes e^{-iH_Bt}\tau_B e^{iH_Bt}U^\dagger\right] \\
&= \mathsf{Tr}_B\left[Ue^{-i(H_S+H_B)t}\rho_S \otimes \tau_B e^{i(H_S+H_B)t}U^\dagger\right] \\
&= \mathsf{Tr}_B\left[e^{-i(H_S+H_B)t}U\rho_S \otimes \tau_B U^\dagger e^{i(H_S+H_B)t}\right] \\
&= \mathcal{R}_t(\mathcal{E}\left(\rho_S\right))
\end{aligned} \tag{5.13}$$

□

We now show that thermal operations commuting with time translations allows us to restrict our attention to quasi-classical states. First, we define the decoherence (dephasing) map

Definition 28 (Decoherence in the energy eigenbasis)**.** We define the following decoherence map, which is equivalent to taking the Haar measure of the group of time translations generated by the systems Hamiltonian

$$\mathcal{D}_S(\rho) = \lim_{T\to\infty}\frac{1}{T}\int_{-T}^{T}\mathcal{R}_t(\rho)dt = \lim_{T\to\infty}\int_{-T}^{T}e^{-iH_St}\left(\rho\right)e^{iH_St}dt \tag{5.14}$$

Taking the limit $\lim_{T\to\infty}\frac{1}{T}\int_{-T}^{T}e^{-i\omega T}dt = 0$ for $\omega \neq 0$, all rotating terms in the density matrix, $e^{-iH_St}\left(\rho\right)e^{iH_St}$ vanish and we are left with the state that is invariant under

time translations. This map therefore destroys all coherences in the energy eigenbasis other than those corresponding to rotating terms for $\omega = 0$, i.e. degenerate sub-spaces in the energy spectrum. Therefore, $\mathcal{D}_S(\rho)$ will commute with the Hamiltonian for all input states ρ and is therefore always a quasi-classical state, as per Definition 27. Whilst in general there could be states with coherences on degenerate eigenstates, we can use an energy preserving unitary on the system alone to rotate these degenerate subspace so that the state is diagonal in the energy eigenbasis. Because we are interested in determining the optimal work cost of transforming quantum states under thermal operations, we can always apply the dephasing map defined in Definition 28 and a energy preserving unitary to the system, without incurring any additional energy cost.

From now on, we restrict ourselves to these 'diagonal states', which can be written as $\rho = \sum_s p(s)|s\rangle\langle s|$ where $\{|s\rangle\}$ are the energy eigenvectors of H_S, $H_s|s\rangle = E_s|s\rangle$. The impact of this is that it is sufficient to consider systems that are diagonal in the energy eigenbasis when determining the ultimate limitations of thermal operations, and do not have to concern ourselves with the intricacies of how coherences impede the set of allowed thermodynamical transformations (see for example [91, 92, 83, 108, 41]).

Lemma 18 allows us to further simplify our treatment of thermal operations. For these states we can characterize thermal operations in terms of a class of classical stochastic maps, rather than having to consider the full set of unitary transformations describing the system-bath interaction.

Theorem 3 (effect of thermal operations on the system)**.** Given a thermal operation $\epsilon(\cdot)$ acting on quasi-classical state ρ, the image $\sigma = \epsilon(\rho)$ is also a quasi-classical state. Furthermore, the energy level populations of the image state $\sigma = \sum_{s'} q(s')|s'\rangle\langle s'|$ are related to population of the initial state $\rho = \sum_s p(s)|s\rangle\langle s|$ by

$$q(s') = \sum_s t(s'|s)p(s) \tag{5.15}$$

where $t(s'|s) = \langle s'|\mathcal{E}\left(|s\rangle\langle s|\right)|s'\rangle$ is a stochastic matrix with the following properties

- *positivity*: $t(s'|s) \geq 0 \ \forall \ s, s'$
- *stochasticity*: $\sum_{s'} t(s'|s) = 1 \ \forall \ s$
- *Gibbs-preserving*: $\sum_i t(s'|s)e^{-\beta E_s}/Z_S = e^{-\beta E_{s'}}/Z_S \ \forall \ s'$

Proof. Definition 28 and Lemma 18 imply that the dephasing map commutes with all thermal operations

$$\mathcal{E}\left(\mathcal{D}_{\mathcal{H}}(\cdot)\right) = \mathcal{D}_{\mathcal{H}}\left(\mathcal{E}(\cdot)\right) \tag{5.16}$$

which implies that the image of a thermal map acting on a quasi-classical state is also a quasi-classical state

$$\begin{aligned} \sigma = \mathcal{E}(\rho) &= \mathcal{E}\left(\mathcal{D}_{\mathcal{H}}(\rho)\right) \\ &= \mathcal{D}_{\mathcal{H}}\left(\mathcal{E}(\rho)\right) \\ &= \mathcal{D}_{\mathcal{H}}(\sigma) \end{aligned} \tag{5.17}$$

We can then decompose the quasi-classical state σ as a convex sum over the energy eigenstates $\sigma = \sum_{s'} q(s')|s'\rangle\langle s'|$, where $q(s') := \langle s'|\sigma|s'\rangle$ and obey $\sum_{s'} q(s') = 1$, $q(s') \geq 0$. By the same argument we can decompose ρ as $\rho = \sum_s p(s)|s\rangle\langle s|$. Using the above definition of $q(s')$ we arrive at

$$
\begin{aligned}
q(s') &:= \langle s'|\sigma|s'\rangle \\
&= q(s') := \langle s'|\mathcal{E}\left(\sum_s p(s)|s\rangle\langle s|\right)|s'\rangle \\
&= \sum_s \langle s'|\mathcal{E}\left(|s\rangle\langle s|\right)|s'\rangle p(s) \\
&= \sum_s t(s'|s)p(s)
\end{aligned}
$$

positivity and stochasticity follow from the properties of positivity and trace-preserving of quantum maps respectively. The Gibbs preserving property is proven in Corollary 2. □

The work storage system

Having reached a simple characterization of the allowed transformations on an isolated system and bath, we now introduce an ancillary system which we refer to as a *weight*. By including a weight, we can allow transformations that change the total energy of the system and bath, with the weight acting as a store for any residual energy shed by the system and bath, or as a battery to provide energy for a system-bath transition that cannot occur spontaneously. We refer to the energy stored in the weight as 'work'. Historically the weight has been thought of as a mechanical system such as a mass on a chord, that raises or lowers against the force of gravity. This is a common definition of a work system and can be abstracted as a work Hamiltonian $H_W = \int_{\mathbb{R}} dx|x\rangle\langle x|$ (where e.g. the orthonormal basis $\{|x\rangle, \forall x \in \mathbb{R}\}$ represents the vertical position of the weight). The following results and definitions are published in [114, 113].

Introducing the weight requires two additional assumptions

Definition 29 (Classical work)**.** Before and after performing the global energy

preserving unitary U on the joint state of system, weight and bath, the state of the weight is measured, obtaining outcomes $|x\rangle$ (initial) and $\{|x + W_i\rangle\}$ with probability

$$p_{W_i} := \langle x + W_i | \mathsf{Tr}_{S,B} \left[U(\rho_S \otimes \rho_B \otimes |x\rangle\langle x|) U^\dagger \right] |x + W_i\rangle$$

(final). In general, we allow the work to be a fluctuating random variable, $\langle W \rangle = \sum_i p_i W_i$.

Definition 30 (Independence on the 'position' on the weight)**.** The global unitary U commutes with the translation operator on the weight

$$[U, \Delta_W] = 0$$

The generator of the translations of the weight Δ_W is canonically conjugated to the position (or potential energy) of the weight, $[H_W, \Delta_W] = i$, where Δ_W is a Hermitian operator and acts as $e^{i\Delta_W \cdot W} |x\rangle_W = |x + W\rangle \ \forall \ x, W \in \mathbb{R}$

The first assumption is that the work we expend or extract in a thermodynamic operation is essentially classical - i.e. measuring the work system destroys any coherences between its eigenstates (the weight has a non-degenerate spectra). This is known in the literature as the *two point measurement scheme*, and is the traditional definition of work in quantum thermodynamics (see [30] for a review). There have been recent attempts to consider non-classical descriptions of work [76, 90, 108]. These approaches have their own limitations, and the problem of finding a definition of 'quantum work' remains an important open question, beyond the scope of this thesis.

The second assumption is required to ensure that we cannot use the weight as a source of free energy. For example, it could be that the weight is inadvertently used as a source of non-uniformity [56]. The second assumption implies that the unitary U cannot reduce the entropy of the joint system-bath combined. As unitaries conserve

total entropy of system, bath and weight, this constraint implies that they cannot decrease the entropy of the weight, which guarantees that the weight is not used as a source of purity. We now show that the above condition ensures this

Lemma 19 (Weight independence)**.** If $[U, \Delta_W] = 0$, then for every state of the weight ρ_W there exists a probability distributions $p(w)$ such that

$$\mathsf{Tr}_W \left[U \rho_S \otimes \rho_B \otimes \rho_W U^\dagger \right] = \int dw\, p(w) u_w \rho_{SB} u_w^\dagger \tag{5.18}$$

where u_w are unitaries on the joint system-bath state. Therefore, as u_w are entropy-preserving, U cannot reduce the entropy of the joint system-bath state. (Theorem 1 from [96])

Proof. By Definition 30, $[U, \mathbb{I}_{S,B} \otimes \Delta_W] = 0$. This implies that we can tensor-decompose the unitary as

$$U = \int dz\, A_z \otimes e^{iz\Delta_W} \tag{5.19}$$

where A_z is a family of operators on $H_S \otimes H_B$. Unitarity implies that

$$\mathbb{I}_{SB} \otimes \mathbb{I}_W = UU^\dagger = \int dz\, dz'\, A_z A_{z'}^\dagger \otimes e^{i(z-z')\Delta_W} \tag{5.20}$$

which implies that

$$\int d_z A_z A_{z+x}^\dagger = \mathbb{I}_{SB} \delta(x) \tag{5.21}$$

where $\delta(x)$ is the Dirac delta function. Let us now characterize the reduced map on system and bath.

$$\begin{aligned}\mathsf{Tr}_W\left[U\rho_{SB}\otimes\rho_W U^\dagger\right] &= \int dz\, dz'\, A_z\rho_{SB}A_{z'}^\dagger \mathsf{Tr}\left[e^{i(z-z')\Delta_W}\rho_W\right] \\ &= \int dz\, dz'\, A_z\rho_{SB}A_{z'}^\dagger \int dk p(k) e^{ik(z-z')}\rho_W \end{aligned} \tag{5.22}$$

where $p(k) := \langle x|\rho_W|k\rangle$ and $|k\rangle$ is an eigenstate of the translation operator $\Delta_k|k\rangle = k|k\rangle\ \forall\ k\in\mathbb{R}$. Define the family of operators $u_k := \int dz\, e^{ikz}$, we can write the above reduced map as

$$\mathsf{Tr}_W\left[U\rho_{SB}\otimes\rho_W U^\dagger\right] = \int dk\, p(k) u_k\rho_{SB}u_k^\dagger \tag{5.23}$$

All that is left is to show that $u_k^\dagger = u_k^{-1}$

$$\begin{aligned} u_k u_k^\dagger &= \int dz\, dz'\, A_z A_{z'} e^{ik(z-z')} \\ &= \int dz\, dx\, A_z A_{z+x} e^{ikx} \\ &= \int dx \delta(x)\mathbb{I}e^{-ikx} = \mathbb{I} \end{aligned}$$

□

5.0.3 Thermal operations with non-constant Hamiltonians

So far, we have only considered thermal maps that preserve the Hamiltonian of the system. Thermal operations are general enough to include the case where the final Hamiltonian of the system H_S' differs from the initial Hamiltonian H_S. This can be shown by including an ancillary qubit X, which acts as a 'switch', as described in [71, 3]. The total Hamiltonian is given by

$$H_{\text{tot}} = H_S \otimes |0\rangle_X\langle 0| + H'_S \otimes |1\rangle_X\langle 1| + H_B + H_W \tag{5.24}$$

We impose that the global energy conserving unitary performs the switching

$$U\left(\rho_{SWB} \otimes |0\rangle_X\langle 0|\right) = \rho'_{SWB} \otimes |1\rangle_X\langle 1| \tag{5.25}$$

for any initial ρ_{SWB}. This implies that

$$U = V \otimes |1\rangle_X\langle 0| + \tilde{V}|0\rangle_X\langle 1| \tag{5.26}$$

Where V and $\tilde{V}$ are unitaries on the system, bath and weight. The condition $[U, H_{\text{tot}}] = 0$ implies that

$$U\left(H_S + H_B + H_W\right) = \left(H'_S + H_B + H_W\right) U \tag{5.27}$$

Note that U does not necessarily commute with $H_{\text{tot}} = H_S + H_B + H_W$, but satisfies (5.27). Therefore the reduced map on system, bath and weight can be written as

$$\mathcal{E}(\rho_{SBW}) = U\rho_{SBW}U^\dagger \tag{5.28}$$

where U satisfies (5.27).

Thermal operations on system and bath

We finish this section by extending our analysis of the thermal operations on the system to a set of analogous constraints for the thermal operations on the system and bath. As we showed in Lemma 19, the reduced map on system and bath is a probabilistic mixture of unitaries. In Lemma 18 we showed that the dephasing

map commutes with the thermal operation reduced map on the system. It is trivial to extend this dephasing map to the system and bath, and (as the trace over bath plays no role in Lemma 18) the result extends to the reduced thermal operation on system and bath. That is, we can restrict our self to considering quasi-classical system and bath states. We now proceed as in Theorem 3. We define the following stochastic matrix, which characterizes the reduced thermal operation on system and bath, where we denote the bath energy eigenstates as $|E\rangle$.

$$t(E'\,s'|E\,s) = \mathsf{Tr}\left[(|s'\rangle\langle s'|\otimes Q_{E'\otimes\mathbb{I}})\,U\left(|s\rangle\langle s|\otimes\frac{Q_E}{\Omega(E)}\otimes\rho_W\right)U^\dagger\right] \tag{5.29}$$

Q_E is the projector onto the (degenerate) eigenspace of H_B that has energy E, and the density of states is $\Omega_E = \mathsf{Tr}\,[Q_E]$. The elements $t(E'\,s'|E\,s)$ can be interpreted as the probability of the joint system and bath microstate undergoing a transition from $|s\rangle\otimes|E\rangle$ to $|s'\rangle\otimes|E'\rangle$ under the action of U. We arrive at the following constraints on the map, analogous to Lemma 3.

Theorem 4 (thermal reduced map on system and bath)**.** The reduced thermal map on system and bath is characterized by the stochastic matrix $t(E'\,s'|E\,s)$ which obey the following necessary and sufficient conditions

1. *stochasticity on the system*

$$\sum_{EE's} p(s)p_b(E)t(E'\,s'|E\,s) = q(s') \quad \forall\, s'$$

2. *law of total probability*

$$\sum_{E's'} t(E'\,s'|E\,s) = 1 \quad \forall\, E, s$$

3. *positivity*

$$t(E'\,s'|E\,s) \geq 0 \quad \forall\, E', s', E, s$$

4. *reversibility*

$$\sum_{Es} t(E'\,s'|E\,s)\Omega(E) = \Omega(E') \quad \forall\, s', E'$$

Proof. Condition 1 is the statement that the reduced map on the system is consistent, in that it maps the system to its final (quasi-classical) state $\sigma = \sum_{s'} q(s')|s'\rangle\langle s'|$, and where we have used the fact that the system and bath are initially in a product state, $p(s, E) = p(s)p_b(E)$. Conditions 2 and 3 are standard conditions for stochastic matrices. Condition 4, as we will see, is the condition that is responsible for the second law of thermodynamics. We assume that the bath has a very large number of microstates of energy E and E'. $\sum_s t(E'\,s'|E\,s)\Omega(E)$ is then equal to the fraction of bath microstates with energy E_b that are mapped by the thermal operation to bath microstates with energy E', and therefore if we sum over all initial energies E we count all of the microstates mapped to energy E'. As the map is reversible, it is a bijection on the set of system-bath states, and therefore the number of microstates with energy E' before and after applying the thermal operation remain the same. □

5.0.4 Work, heat and transformations

We have now defined the basic set-up in which we will analyze a variety of thermodynamic protocols - a thermal bath, an (athermal) system, and a work storage system or 'weight' that tallies the change in energy of the system and bath at the end of their interaction. In our work, we focus exclusively on *work extraction* and *work of formation* protocols. In the former we either allow heat to flow from the bath to the system, thermalising (or partially thermalising) the system, and extracting work into the weight along the way, whereas in the latter the interaction consumes work stored in the weight, in order to move the system further from thermal equilibrium. Common examples of protocols falling into these categories include cooling (work of formation), and isothermal expansion (work extraction). For now, we will consider the case where we have a single thermal bath at some fixed temperature β, but later

we will see that single-bath protocols can be used as steps in multi-bath protocols, such as constructing Carnot engines.

We will consider these protocols under additional constraints to the work. In general, the work will be a a fluctuating random variable W. If it is allowed to freely fluctuate, we could denote this as $W^{(\infty)}$, indicating that we allow fluctuations of arbitrary size. In many applications, there is a desire to bound the size of these fluctuations - for example if we are working with small or fragile machines that cannot accommodate arbitrary energetic fluctuations, or if we are interested in determining the deterministic work content of a system (the *single-shot* work, where the work distribution is guaranteed to be perfectly concentrated [71, 2]). If we bound the fluctuations in the work to be smaller than some threshold $|W_i - \langle W\rangle| \leq C$, we denote this as $W^{(c)}$. If we demand that the work is deterministic (i.e. the work distribution is non-zero at a single value), we denote this as $W^{(0)}$, and if we allow for some very small ϵ fluctuation, we denote this as $W^{(\epsilon)}$ (note, this is often referred to as an ϵ probability of failure, but this is nothing more than allowing the work to be of the form $\langle W\rangle = (1-\epsilon)W + \epsilon \times 0$, and thus have a single fluctuation with $W_i = 0$). Regardless of the specific constraints we place on the work, we can define the following optimal work measures, which are tied to specific protocols of interest

Definition 31 (Work of transition)**.** Given two quasi-classical states ρ_S, σ_S, the maximal work gain associated with the transition $\rho_S \to \sigma_S$ under a thermal operation is $W_{\rho\to\sigma}$, defined as

$$\begin{aligned} W_{\rho\to\sigma} = \max_{\mathcal{E}} \{\langle W\rangle = \mathsf{Tr}_W \left[H_W \mathsf{Tr}_{S,B}\left[\mathcal{E}\left(\rho_S \otimes \rho_B \otimes |0\rangle_W\langle 0|\right)\right]\right] \\ \text{s.t.} \quad \mathsf{Tr}_{B,W}\left[\mathcal{E}\left(\rho_S \otimes \rho_B \otimes |0\rangle_W\langle 0|\right)\right] = \sigma \end{aligned} \tag{5.30}$$

In other words, it is the minimal work done on the system S by a thermal operation that achieves the transformation on the system $\rho_S \to \sigma_S$.

Note that, in general, the system will end up in a correlated state with the weight (i.e. separable, due to measuring the work system, but not in a product state). In the case that work is deterministic, the final system-weight separable state must be a product state, and the above optimization simplifies to finding the maximal final weight energy W.

Definition 32 (Work content)**.** The work content of a system is the maximal work extracted by a thermal operation, without constraining the final state of the system

$$W_\rho = \max_{\mathcal{E}} \left\{ \langle W \rangle = \mathsf{Tr}_W \left[H_W \mathsf{Tr}_{S,B} \left[\mathcal{E} \left(\rho_S \otimes \rho_B \otimes |0\rangle_W \langle 0| \right) \right] \right] \right\} \tag{5.31}$$

As we will show, if work is allowed to fluctuate freely, any thermal operation that extracts the full work content of the system leaves it in the thermal state w.r.t the inverse temperature of the bath. Therefore (in this case) we can think of optimal work extraction as analogous to thermalizing the system.

Definition 33 (work of formation)**.** The work of formation is the minimal work cost of forming the state ρ_S from its thermal state w.r.t the bath, $\tau_S = Z^{-1} e^{-\beta H_S}$ using thermal operations

$$\begin{aligned} W_{\tau \to \rho} = \max_{\mathcal{E}} \left\{ \langle W \rangle = \mathsf{Tr}_W \left[H_W \mathsf{Tr}_{S,B} \left[\mathcal{E} \left(\tau_S \otimes \rho_B \otimes |0\rangle_W \langle 0| \right) \right] \right] \right. \\ \left. \text{s.t.} \quad \mathsf{Tr}_{B,W} \left[\mathcal{E} \left(\tau_S \otimes \rho_B \otimes |0\rangle_W \langle 0| \right) \right] = \rho \right\} \end{aligned} \tag{5.32}$$

where $\tau_S = Z^{-1} e^{-\beta H'_S}$

5.0.5 Landauer erasure: a simple illustration

Before we move on to deriving the fundamental bounds on the work cost of thermal operations, it will be useful to illustrate how our framework can be used with a simple example. Perhaps the simplest and often studied protocol is Landauer erasure [84, 20]. In this protocol a 2 dimensional system in an unknown state (i.e. the maximally mixed state $\rho = 1/2|0\rangle\langle 0| + 1/2|1\rangle\langle 1|$) is 'erased', by resetting the state of the system to a fixed pure state (e.g. $\sigma = |0\rangle\langle 0|$). In this simple case, we see what the optimal thermal operations (defined by $\{t(E', s'|E, s)\}$) look like quite easily. A thermal operation is essentially a probabilistic permutation of the joint states of system and bath. In figure 5.1 we show the joint depict the joint system-bath states by drawing the bath microstates (growing by $\Omega(E) \propto e^{\beta E}$ for an infinite bath), and stratifying them along the two sub-spaces of the system, $|0\rangle$ and $|1\rangle$

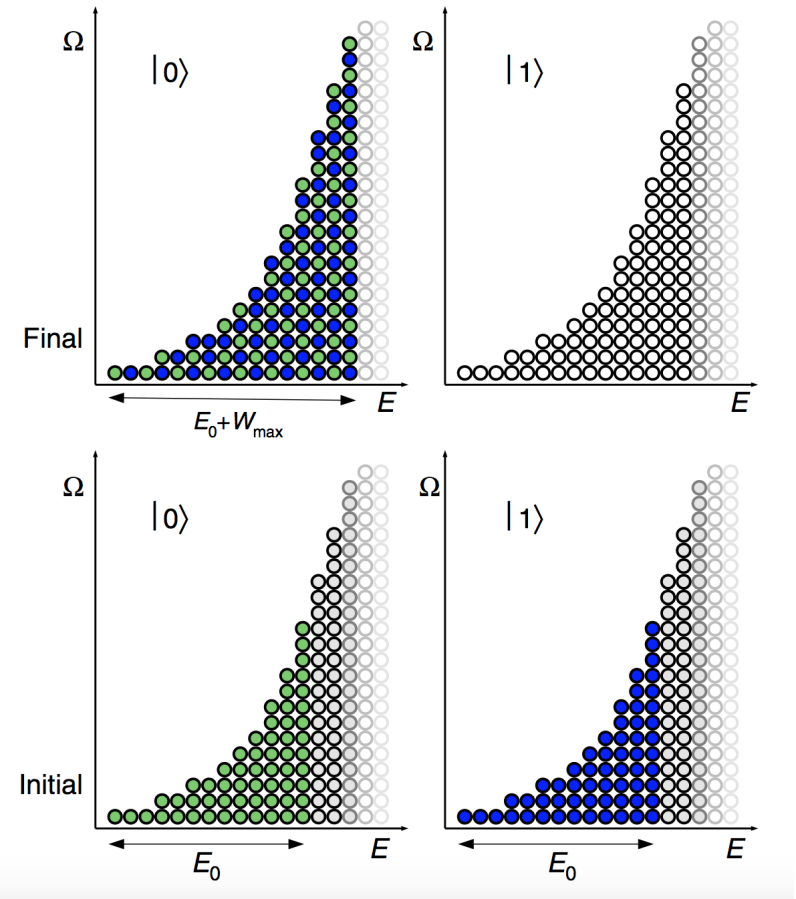


Figure 5.1: figure from [95], reprinted with permission.

In figure 5.1, the initial joint microstates of system and bath are shown, with the states with energy $E \leq E_0$ coloured green in the subspace spanned by $|0\rangle$, and blue in the subspace spanned by $|1\rangle$. The work cost of erasure can clearly be seen. As the thermal operation is reversible (a bijection on the infinite-dimensional joint state space), the final joint state has to accommodate all of the initial microstates, and we have to map the initial joint system-bath states to bath states of higher energy to do so. The thermal operation maps all states to the subspace spanned by $|0\rangle$, i.e. the probability of finding the reduced state of the system in $|0\rangle$ is 1. In other words, $t(E', s'|E, s) = t(E'|E, s)\delta(s')^0$.

Consider the case where we want to perform erasure with a deterministic work cost (as we will see later, for erasure this is also the optimal work cost if we do not constrain fluctuations). Then we require that $t(E', s'|E, s) \neq 0 \implies E' = E_b + W$. Consider the energy subspace of the final state of the bath $H_B(x_B) = E$. This must contain the image states from bath energy levels $E - W$, as shown in figure 5.1. As there are two such energy levels (due to the two system sub-spaces present in the preimage), reversibility implies that $\Omega(E) = 2\Omega(E - W)$ which implies $W = \beta^{-1} \log 2$, which is Landauer's bound. This is not the most general case, as the work as been constrained to be deterministic. However, the proofs in the following section follow the same logic - using reversibility and knowledge of the density of states of the bath to determine the optimal work.

5.1 Fluctuating work

We can now derive some of the standard results of thermodynamics within our framework. These are the results that we will then go on to extend in later sections. Firstly, we deal with the case that work stored in the weight is allowed to fluctuate arbitrarily. Without placing any constraints on the weight, the only resulting constraint on what transformations can be achieved between quasi-classical states is the

corresponding work cost. In other words, the only constraint is that there must be enough work stored in the weight to accommodate the work cost of transition $W_{\rho\to\sigma}$. It is often also said that a system in contact with a thermal bath can only undergo a transition spontaneously if doing so generates surplus work, $W_{\rho\to\sigma} \geq 0$. Using Theorem 4 we now quantify the *optimal work* achievable in a given state transition.

Lemma 20 (Work of transition with infinite bath and unbounded fluctuations)**.** The work extracted by a thermal operation $t(E'\,s'|E\,s)$ is upper bounded by $W_{\rho\to\sigma}$ where

$$W_{\rho\to\sigma} = F(\rho) - F(\sigma) - \frac{1}{\beta}\mathcal{I}(\mathcal{E}(\rho_S \otimes \rho_B)) \tag{5.33}$$

where

$$F(\psi) = U(\psi) - \frac{1}{\beta}S(\rho) \tag{5.34}$$

where $S(\rho)$ is the Shannon entropy of the quasi-classical state $\rho = \sum_i p_i|i\rangle\langle i|$, $S(\rho) = -\sum_i p_i \log p_i$, $U(\rho) = \mathsf{Tr}\left[H_\rho \rho\right]$ and $\mathcal{I}(S:B)$ is the mutual information between the final states of the system and bath. Furthermore, in the case that the thermal bath is infinite, there exists thermal operation that achieve upper bound with zero mutual information

$$W_{\rho\to\sigma} = F(\rho) - F(\sigma) \tag{5.35}$$

Proof. The expected work extracted by a thermal operation characterized by $t(E'\,s'|E\,s)$ is given by the expected reduction in the energy of the system and bath

$$\langle W\rangle = \sum_{EE'ss'} p(s)p_b(E)t(E'\,s'|E\,s)\left[(E-E') + (\epsilon_s - \epsilon_{s'})\right] \tag{5.36}$$

where $H_S|s\rangle = \epsilon_s|s\rangle$ and $p_b(E) = Z_b^{-1}\Omega(E)e^{-\beta E}$ is the probability of finding the bath in the energy subspace E. For any $t(E'\,s'|E\,s)$ we can rewrite this expression

for the expectation value of the work in terms of the following variables that are independent of the initial bath energy E

$$R(E', s'|s) := \sum_E p_b(E) t(E'\, s'|E\, s) \tag{5.37}$$

$$Q(E', s'|s) := \frac{\sum_E p_b(E) t(E'\, s'|E\, s)(E - E')}{R(E', s'|s)} \tag{5.38}$$

Which allow us to re-express the $\langle W \rangle$ associated with $t(E'\, s'|E\, s)$ as

$$\langle W \rangle = \sum_{E' s s'} p(S) R(E' s'|s) Q(E', s'|s) + \Delta U(S) \tag{5.39}$$

Where $\Delta U(S) = \sum_{E E' s s'} p(s) p_b(E) t(E'\, s'|E\, s)(\epsilon_s - \epsilon_{s'}) = \sum_{s'} p(S')(\epsilon_s - \epsilon_{s'})$ is the change in the internal energy of the system under the map (where we have used constraints 1 and 2 of theorem 4. Next we rephrase condition 4 (reversibility) of theorem 4 in terms of these new map variables

$$\begin{aligned}
1 &= \sum_{Es} t(E'\, s'|E\, s) \frac{\Omega(E)}{\Omega(E')} \\
&= \sum_{Es} t(E'\, s'|E\, s) \frac{p_b(E)}{p_b(E')} e^{-\beta(E - E')} \\
&= \sum_{Es} \frac{R(E' s'|s)}{R(E' s'|s)} t(E'\, s'|E\, s) \frac{p_b(E)}{p_b(E')} e^{-\beta(E - E')} \\
&\geq \sum_{s} \frac{R(E' s'|s)}{p_b(E')} \exp\left[\beta \frac{\sum_E p_b(E) t(E' s'|Es)(E - E')}{R(E' s'|s)}\right] \\
&= \sum_{s} R(E' s'|s) \frac{e^{\beta Q(E' s' s)}}{p_b(E')}
\end{aligned} \tag{5.40}$$

We can now determine the optimal $\langle W \rangle$ by maximizing (5.39) w.r.t this constraint along with the others in Theorem 4. As we will see, it will be sufficient to construct the Lagrangian with the reversibility constraint alone and then apply the

stochasticity constraints to the optima of this Lagrangian

$$\mathcal{L} = \sum_{E's s'} p(S)R(E's'|s)Q(E',s'|s)+\Delta U(S)+\sum_{E's'} \lambda_{E',s}\left(p_b(E') - \sum_s R(E's'|s)e^{\beta Q(E's's)}\right) \tag{5.41}$$

where $\lambda_{E',s}$ are the Lagrange multipliers for the the constraints

$p_b(E') \geq \sum_s R(E's'|s)\mathsf{exp}\left(\beta Q(E's's)\right)$ from (5.40). The solution to $\partial_Q \mathcal{L} = 0$ yields

$$Q(E',s',s) = \frac{1}{\beta}\log\left(\frac{p(s)}{\lambda_{E's'}\beta}\right) \tag{5.42}$$

substituting this back into the Lagrangian and taking the derivative w.r.t the Lagrange multiplier $\lambda_{E's'}$ and equating to zero gives

$$\frac{1}{\lambda_{E's'}\beta} = \frac{p_b(E')}{\sum_s P(s)R(E's'|s)} = \frac{p_b(E')}{P(E's')}\ , \tag{5.43}$$

where for the last equality we used the definition of R. This gives the optimal Q

$$Q(E's's) = \frac{1}{\beta}\log\left(\frac{P(s)p_b(E')}{P(E's')}\right)\ . \tag{5.44}$$

And thus the optimal work

$$\langle W\rangle \leq \Delta U + \frac{1}{\beta}\sum_{E's s'} P(s)R(E's'|s)\log\left(\frac{P(s)p_b(E')}{P(E's')}\right)\ , \tag{5.45}$$

which can also be written as

$$\begin{aligned}\langle W\rangle \ &\leq\ \Delta U - \frac{1}{\beta}\Delta S - \frac{1}{\beta}\sum_{E's s'} P(s)R(E's'|s)\log\left(\frac{P(s)}{P(s')}\right) &(5.46)\\ &+\ \frac{1}{\beta}\sum_{E's s'} P(s)R(E's'|s)\log\left(\frac{P(s)p_b(E')}{P(E's')}\right)\ , &(5.47)\end{aligned}$$

where $\Delta S = S(\rho) - S(\sigma)$ which, under constraints 1-2 of Theorem 4 simplifies to

$$\langle W\rangle \leq -\Delta F - \frac{1}{\beta}\sum_{E's'} P(E's')\log\left(\frac{P(E's')}{p_b(E')P(s')}\right) = -\Delta F - \frac{1}{\beta}\mathcal{I}(\mathcal{E}(\rho_S\otimes\rho_B)) \quad (5.48)$$

□

This bounds gives an absolute upper limit on the work that can be extracted by a thermal operation. Note that the upper bound is just the standard second law, $\langle W\rangle = -\Delta F$, but with a correction coming from the correlations that remain between system and bath after they have interacted. Clearly, the system and bath ending up completely uncorrelated is a very idealized situation, and we will show is only possible in the limit that the thermal bath becomes infinite. Note that we made no assumption on the structure of the thermal bath, other than the convexity of its density of states. As we will see, this means that our bound applies to finite bath models, and the finite bath corrections that we later derive are essentially bounds on the minimal correlation achievable between system and bath when extracting some expected work.

Clearly, by the positivity of quantum and classical mutual information, there is a hard upper-bound on the work of $-\Delta F$. We now show that for infinite baths this upper bound is attainable.

Lemma 21. For infinite baths, where $\Omega(E) \propto e^{\beta E}$, there exists optimal thermal operations that saturate the inequality $\langle W\rangle \geq -\Delta F$. All such thermal operations are of the form

$$t(E's'|Es) = f(E's's)\delta(E - E' - f_{s'} + f_s)$$

where $f(E's's)$ obeys

$$\sum_{s'} f(E - f_{s'} + f_s, s', s) = 1 \quad \forall\ s \tag{5.49}$$

$$\sum_{s} P(s) f(E's's) = P(s') \quad \forall\ s' \tag{5.50}$$

$$0 \leq f(E's's) \leq 1 \quad \forall\ E's's \tag{5.51}$$

where $f_s = -\beta^{-1} \log P(s)$ is the fine-grained entropy.

Proof. In order to achieve $\langle W \rangle = -\Delta F$, a thermal operation $t(E's'|Es)$ must saturate the upper bound derived in Lemma 20 with zero mutual information. First note that in Lemma 20 the optimal $Q(E's's)$ are (uniquely)

$$Q(E's's) = \frac{1}{\beta} \log \left(\frac{P(s) p_b(E')}{p(E's')} \right) . \tag{5.52}$$

As the mutual information is a distance measure for probability distributions the upper bound reduces to $-\Delta F$ if and only if $P(E's') = p_b(E')P(s')$ and therefore, for all optimal thermal operations

$$Q(E's's) = \frac{1}{\beta} \log \left(\frac{P(s)}{P(s')} \right) = f_s - f_{s'} . \tag{5.53}$$

where we define the fine-grained entropies as $f_s := -\log p(s)$ (also referred to as the 'surprise' or 'surprisal'). Finally, note that the thermal operation only construct an optimal Q if the reversibility constraint (5.40) is saturated. This requires that

$$\sum_{Es} \frac{R(E's'|s)}{p_b(E')} t(E's'|Es) \frac{p_b(E)}{R(E's'|s)} e^{\beta(E-E')} = \sum_{s} \frac{R(E's'|s)}{p_b(E')} \exp\left[\beta \frac{\sum_E p_b(E) t(E's'|E_n s)(E - E')}{R(E's'|s)} \right] \tag{5.54}$$

Therefore the sum over E in the exponent must have only a single term by convexity, and therefore $t(E's'|Es)$ must be a delta function killing the sum over E. This,

combined with the definition of $Q(E's's)$ implies that $t(E's'|Es)$ is of the form

$$t(E's'|Es) = f(E's's)\delta(E - E' - f_{s'} + f_s) \tag{5.55}$$

where $f(E's's)$ is a function that, by substituting (29) into Constraints 1-3 in Theorem 4 yields

$$\begin{aligned} \sum_{s'} f(E - f_{s'} + f_s, s', s) &= 1 \quad \forall\ s & (5.56)\\ \sum_{s} P(s) f(E's's) p_b(E' + f_{s'} - f_s) &= P(s') p_b(E') \quad \forall\ s', E' & (5.57)\\ 0 \leq f(E's's) &\leq 1 \quad \forall\ E's's & (5.58) \end{aligned}$$

IN the following we denote $\gamma := \beta^2/C$. In the limit $\gamma \to 0$ the distribution $p_b(E)$ tends to a constant. Hence, equality (5.57) becomes (5.50). □

Lemma 21 shines a light on why we can miraculously ignore the correlations between the system and bath when working in the infinite bath limit, and just consider the reduced map on system. Whats-more, we have managed to phrase the problem of computing the optimal work as a convex optimization problem, which we are now free to throw additional constraints into. The obvious assumptions to attack with additional constraints are the infinite size of the bath, and the infinite fluctuations of the work. In the following section we use our framework to derive the main results of *single-shot work extraction* [2, 71], where work is constrained to be deterministic. We start down this path by deriving the simplest constrained work - the *single-shot* work where the work fluctuations are constrained to be zero.

5.2 Deterministic work extraction

Single-shot work, introduced in [2] and [71], is simply the optimal work associated with a given thermodynamic process (e.g. minimal work of formation, maximal

work content) when we constrain the work from being a random variable to being deterministic. This approach can then be extended to allow for a small ϵ probability of failure, where we extract work W with probability $1-\epsilon$ or our protocol 'fails' with probability ϵ, as in [2, 71, 63, 118]. Ultimately, this is allowing the work to fluctuate, with some small probability but perhaps with very large fluctuation size. In this section we derive the single-shot work in the limit $\epsilon \to 0$, i.e. truly deterministic work, and then in the proceeding section generalize this to allow for (constrained) fluctuations in a more natural way than allowing for a probability of failure.

We will focus on two protocols that allow the single-shot work to be written in a closed form, the work of formation and work content (as we will see later, constraining the size of work fluctuations makes work of transition difficult to calculate). For now, we will simply present expressions for the single-shot work, although later we will derive them simply from our own formulae.

Definition 34 (Single-shot work content)**.** The optimal deterministic work that can be extracted from quasi-classical state ρ in contact with an infinite thermal bath of inverse temperature β is given by

$$W_\rho^{(0)} = \frac{1}{\beta}\log Z_S - \frac{1}{\beta}\log\left(\sum_s p(s)^0 e^{-\beta E_s}\right) \tag{5.59}$$

for $\rho = \sum_s p(s)|s\rangle\langle s|$, $Z_S = \sum_s e^{-\beta E_s}$ and $H_s|s\rangle = E_s|s\rangle$, and $p(s)^0 = 1$ if $p(s) \geq 0$, else 0.

Definition 35 (Single-shot work of formation)**.** The minimal work required to transform thermal state τ to quasi-classical state ρ is given by

$$W_F^{(0)} = \min_s \left\{-\frac{1}{\beta}\log\left(p(s)e^{-\beta E_s}Z_S\right)\right\} \tag{5.60}$$

where $\rho = \sum_s p(s)|E_s\rangle\langle E_s|$ and $H_s|s\rangle = E_s|s\rangle$.

The single-shot work content is very prohibitive, in that no work can be extracted

deterministically from states of full rank, regardless of how athermal the initial state is. Likewise, the work of formation is highly restrictive. Whereas the fluctuating work of formation is given by the difference in the Helmholtz free energy between the thermal and target state, which can be expressed as

$$W_F^{(\infty)} = F(\rho) - F(\tau) = -\frac{1}{\beta}\sum_s p(s)\log\left(p(s)e^{-\beta E_S}\right) - \frac{1}{\beta}\log Z_S \tag{5.61}$$

the single shot work of formation takes the most negative (i.e. highest work cost) element of this convex sum (as we will see, this is equivalent to the worst case work fluctuation in $W_F^{(\infty)}$). In all states (bar the thermal state and pure states) $W_F^{(0)} < W_F^{(\infty)} = -W_\rho^{(\infty)} < -W_\rho^{(0)}$. This discrepancy between the work cost and work content of states in the single-shot regime results in thermodynamic irreversibility when transforming between states. Furthermore, the set of allowed thermodynamic transformations in the single-shot regime are severely restricted. In this regime, it is possible for a state to undergo a transition deterministically (and without supplying work) on the condition that a family of 'second laws' are satisfied [24]. Some transitions can only be achieved by supplying work in both forward and backward directions, resulting in a partial order on the set of states with respect to deterministic work. This is in stark contrast to when we allow work to fluctuate freely, whereby all states can be inter-converted in a thermodynamically reversible manner, mediated by a single second law, $\Delta W \leq -\Delta F$

5.3 Towards a unified free energy

We can now present one of our main contributions. We derive a free energy that interpolates between the two cases outlines in the previous section - deterministic and freely fluctuating work. The work presented in this section have been published as [114]. To motivate this result beyond a pleasing unification of the two existing schemes, we prove the following result concerning thermodynamically reversibly

thermal operations (that is, thermal operations for which $W_{\rho\to\sigma} = -W_{\sigma\to\rho} = -\Delta F$.

Theorem 5 (Divergence of fluctuations in thermodynamically reversible protocols)**.** There exists a thermodynamically reversible thermal map ($W_{\rho\to\sigma} = -W_{\sigma\to\rho} = -\Delta F$) with work fluctuations bounded by $|\langle W\rangle - w| \leq c$ if an only if

$$e^{\beta(\Delta F - c)} \leq \frac{p(s)e^{\beta E_s}}{p(s')e^{\beta E_{s'}}} \leq e^{\beta(\Delta F + c)} \quad \forall s, s' \tag{5.62}$$

where s, s' enumerate the energy eigenstates of the initial / final system Hamiltonians.

First, we derive a simplified set of thermal operations that have minimal fluctuations.

Lemma 22 (Optimal thermal operations have minimal work fluctuations)**.** For any thermal operation with reduced map on system and bath given by $t(E's'|Es)$, we can define another thermal operation

$$\tilde{t}(E's'|E,s) := t(s'|s)\delta\left(E' - E - Q(s',s)\right) \tag{5.63}$$

$$Q(s',s) := \frac{\sum_{E'E} p_b(E)t(E's'|Es)(E'-E)}{t(s'|s)} \tag{5.64}$$

where $t('s|s) = \sum_{E'E} p_b(E_b)t(E's'|Es)$ is the reduced map on the system and $Q(s,s')$ is the heat extracted from the bath under the state transition $s \to s'$. This map satisfies all constraints of Theorem 4, and satisfies

$$\langle\tilde{W}\rangle = \langle W\rangle \tag{5.65}$$

$$\max\left|\langle\tilde{W}\rangle - w\right| \leq \max|\langle W\rangle - w| \tag{5.66}$$

where w are the values of the given work distribution. I.e. the two thermal operation have identical expected work, but $\tilde{t}(E's'|E,s)$ has smaller work fluctuations. The

reversibility constraint on $t(E's'|Es)$ becomes

$$1 = \sum_s t(s'|s) e^{\beta Q(s,s')} \tag{5.67}$$

Proof. First, we show $\tilde{t}(E's'|E,s)$ is a legitimate thermal operation. Constraints 2 and 3 of Theorem 4 are immediately satisfied by $t(s'|s)$. To see that constraint 1 is satisfied, we substitute in our definition of $\tilde{t}(E's'|E,s)$, giving

$$\begin{aligned}\sum_{EE's} p(s)p_b(E)\tilde{t}(E'\,s'|E\,s) &= \sum_{E's} p(s)p_b(E'+Q(s',s))t(s'|s) \\ &= \sum_s p(s)t(s'|s) \\ &= q(s') \quad \forall s'\end{aligned} \tag{5.68}$$

where we have used the fact that, for infinite baths, $\Omega(E)p_b(E) = p(E) = p(E' + Q(s',s)) = \text{const}$. Finally, for infinite volume heat baths constraint 4 (reversibility) reduces to

$$\begin{aligned}\Omega(E') &= \sum_{Es's} \tilde{t}(E'\,s'|E\,s)\Omega(E) \\ &= \sum_{s's} t(s'|s)\Omega(E'+Q(s,s')) \\ &= \sum_{s's} t(s'|s)e^{\beta Q(s,s')}\Omega(E')\end{aligned} \tag{5.69}$$

where the $\Omega(E')$ terms cancel, yielding the simplified reversibility constraint.

finally, (5.65) follows trivially by substituting $\tilde{t}(E's'|E,s)$ into the expression for work (5.36). For a given system transition $s \to s'$, the work value associated with $\tilde{t}(E's'|E,s)$ is $Q(s',s) + E_{s'} - E_s$. This is a convex combination of the work values

for $t(E's'|E,s)$ for fixed s, s', and therefore (as $\langle\tilde{W}\rangle = \langle W\rangle$) have fluctuations from the mean that are bounded by the fluctuations associated with $t(E's'|E,s)$.

□

Proof of Theorem 5. We proceed by deriving the work fluctuations associated with the optimal thermal operation, as defined in Lemma 22. As we have shown that these thermal operations have both optimal work (and are therefore thermodynamically reversible) and minimal work fluctuations, their work fluctuations give a tight bound on the minimal thermal operations that achieve thermodynamic reversibility. For our simplified maps, the constraints of Theorem 4 simplify to a set of constraints on the reduced map on the system

1. *stochasticity*

$$q(s') = \sum_s p(s)t(s'|s) \quad \forall \quad s' \tag{5.70}$$

2. *law of total probability*

$$1 = \sum_{s'} t(s'|s) \quad \forall\, s \tag{5.71}$$

3. *positivity*

$$t(s'|s) \geq 0 \quad \forall\, s, s' \tag{5.72}$$

4. *reversibility*

$$1 = \sum_s t(s'|s) e^{\beta(w(s,s')+E_{s'}-E_s)} \tag{5.73}$$

Where we have substituted $w(s,s') = Q(s,s') - E_{s'} + E_s$ which is the work associated with the transition $s \to s'$. It will be useful to relax the reversibility equality to an inequality

$$1 \geq \sum_s t(s'|s)\, e^{\beta(E_{s'}-E_s+w(s',s))} \tag{5.74}$$

where saturation of the inequality implies that the map is an allowed (thermal) operation[4]. In all future calculations we make use of this relaxed constraint and show

[4]This follows from

that our solutions saturate the inequality.

The average work associated with the optimal map achieving the quasi-classical state transition $\rho \to \sigma$ is given by

$$W = \sum_{ss'} p(s)t(s'|s)w(s',s) \tag{5.75}$$

It turns out to be sufficient to optimize this under the reversibility constraint (5.74). The Lagrangian is

$$\mathcal{L} = \sum_{ss'} p(s)t(s'|s)w(s',s) + \sum_{s'} \lambda_{s'} \left(1 - \sum_{ss'} t(s'|s) e^{\beta(w(s',s)+E_{s'}-E_s)} \right) \tag{5.76}$$

We will denote $\Delta E(s,s') = E_{s'} - E_s$. Maximizing with respect to $w(s',s)$ gives

$$w(s',s) = \frac{1}{\beta} \log \left(\frac{p(s)e^{\beta \Delta E(ss')}}{\lambda_{s'}\beta} \right) \tag{5.77}$$

Extremizing with respect to $\lambda_{s'}$ and applying (5.70) and (5.71) gives

$$\lambda_{s'} = \frac{q(s')}{\beta} \tag{5.78}$$

Substituting this into (5.77) gives the optimal work

$$\sum_{ss'} p(s)t(s'|s) \frac{1}{\beta} \log \left(\frac{p(s)e^{\beta \Delta E(s,s')}}{p(s')} \right) = F(\rho) - F(\rho') \tag{5.79}$$

Notice that the result is independent of our choice of $t(s'|s)$, i.e. any map that takes us from $\rho \to \sigma$ gives the same optimal work. This is simply a statement that the free energy is a state variable (i.e. is path independent). The work $W^{(\infty)}$ is the average of the work values, given by

$$w(s',s) = \frac{1}{\beta} \log \left(\frac{p(s)e^{\beta \Delta E(s,s')}}{q(s')} \right) \tag{5.80}$$

Note work values with $p(s) = 0$ are set to zero (they have zero probability of occurring in the work distribution). Note that the fluctuations diverge as the initial / final state moves further from equilibrium. It is easy to check that substituting our solutions for $w(s', s)$ into (5.73) saturates the inequality, therefore this map is achievable with thermal operations.

Lemma 22 shows that no thermal map can achieve this $\langle W \rangle$ with smaller worst case fluctuations. Having explicitly calculated the work values, we have derived necessary and sufficient conditions for a thermal map to exist that achieves the optimal work (i.e. the thermodynamically reversible work) given that the fluctuations are constrained $|\langle W \rangle - w| \leq c$. (5.62) then follows immediately by constraining (5.80) to lie in the interval $\Delta F \pm c$. $\square$

Note that for any finite c there exist states such that Theorem 5 cannot be satisfied. These bounds have strong consequences for the minimal fluctuations that can be achieved with a thermodynamically reversible protocol. For example

$$\lim_{p(s)\to 0} \log \frac{p(s)e^{\beta E_s}}{q(s')e^{\beta E_{s'}}} = -\lim_{q(s')\to 0} \log \frac{p(s)e^{\beta E_s}}{q(s')e^{\beta E_{s'}}} = -\infty \tag{5.81}$$

Therefore as an energy level occupation probability of an eigenstate of either the initial or final quasi-classical state tends to zero the work fluctuation associated with reversibly transitioning to or from this energy level diverges, negatively for work extraction and positively for state formation. In either case we require $c \to \infty$ in order to satisfy inequalities Theorem 5 for all state transformations.

Theorem 5 tells us that the further from equilibrium the initial or final states are, the larger the work fluctuations will be in thermodynamically reversible protocols. If we desire smaller fluctuations in work, we must incur an intrinsic work cost to do so. Note that cooling a system close to its ground state is an example of transitioning from the thermal state to a far from equilibrium state. Defining the ground state

as $|\min\rangle = \min_{E_s}\{|s\rangle \,|\, E_s = \mathsf{Tr}\left[\sigma_s H'_S\right]\}$, then reversible cooling is the transformation $\rho \to |\min\rangle\langle\min|$, which amounts to *purifying* the state. The work fluctuations associated with any intermediate state of the system ρ' with $\mathsf{Tr}\left[\rho'|\min\rangle\langle\min|\right] = 1-\epsilon$ will have fluctuations that diverge like $\log \epsilon$. Similarly, if we want to extract work form a far from equilibrium state using a thermodynamically reversible transformation we encounter the same divergence in fluctuations. The fluctuations can diverge even if the average work remains small (for example, if the system is a qubit with trivial Hamiltonian then $W \leq \beta^{-1}\log 2$). These divergences have been previously noted in the study of absolute irreversibility [52, 99, 98].

Previous criticisms of the inadequacy of the standard free energy have focused on the definition of work [42, 2, 71]. Theorem 5 provides an additional criticism to the standard Helmholtz free energy, in that necessitates divergent work fluctuations, which are clearly an un-physical prediction. We now turn our attention to how we can remove these fluctuations in a principled way.

Theorem 6 (c-bounded work)**.** From [114]. The c-bounded work is defined as the maximum expected work associated with $\rho \to \sigma$ s.t. $|\langle W\rangle - w| \leq c$ where w are the values of the work distribution. For a system S in contact with an infinite thermal bath, the c-bounded work content and work of formation of quasi-classical state ρ are

$$W_\rho^{(c)} = \frac{1}{\beta}\log Z'_S - \frac{1}{\beta}\log \sum_s p(s)^0 e^{-\beta(E_s - \theta_s^{(c)})} \tag{5.82}$$

$$W_F^{(c)} = \frac{1}{X_{\mathrm{u}}}\left[\sum_{s\in\mathcal{X}_{\mathrm{u}}} \frac{p(s)}{\beta}\log\Big(p(s)e^{\beta E'_s}Z\Big) + c\,(1-X_{\mathrm{u}})\right] \tag{5.83}$$

where

$$\theta_s^{(c)} = \begin{cases} \frac{1}{\beta}\log\left(p(s)e^{\beta E_s}\right) - \nu\,,\ sX_u \\ +c\,,\ s \in X_+ \\ -c\,,\ s \in X_i \end{cases} \tag{5.84}$$

where X_u, $X_\pm$ are partitions of the set of system eigenstates by a criterion described in Appendix 1.

Proof. See Appendix 1 □

First we describe the terms in (5.82). $Z' = \sum_{s'} e^{-\beta E_{s'}}$ is the partition function of the final Hamiltonian $H'_{\rm S}$. The second term can be viewed as a generalization of the min free energy (Definition 34) that allows for fluctuations in work $\theta_s^{(c)}$. The only difference to the min free energy is the term $e^{\beta\theta_s^{(c)}}$ included in the summation. $\theta_s^{(c)}$ is the fluctuation of the work value from the c-bounded average $W^{(c)}(\rho)$ given that the system was initially in state $|s\rangle$. In order to find the fluctuations associated with the optimal c-bounded work extraction protocol we must partition the energy levels into three disjoint subsets $\{1, 2, \ldots, d\} = \mathcal{X}_{\rm u} \cup \mathcal{X}_+ \cup \mathcal{X}_-$ (where d is the dimension of the quasi-classical state,) representing the energy levels with positive $\mathcal{X}_+$, negative $\mathcal{X}_-$ and 'unbounded' $\mathcal{X}_{\rm u}$ fluctuations[5]. We also define

$$X_{\rm u} = \sum_{s\in\mathcal{X}_{\rm u}} p(s)\ , \tag{5.85}$$

$$X_\pm = \sum_{s\in\mathcal{X}_\pm} p(s)\ . \tag{5.86}$$

The algorithm for determining the partition $\mathcal{X}_{\rm u} \cup \mathcal{X}_+ \cup \mathcal{X}_-$, which requires the checking of at most $d-1$ inequalities, is described Appendix 1.2. Once we have determined the partition, the fluctuations are given by

[5]These fluctuations are still bounded. However their bounds are not active - the optimal solution is the same as that achieved without bounds on these fluctuations

$$\theta_s^{(c)} = \begin{cases} \frac{1}{\beta}\log(p(s)e^{\beta E_s}) - \nu, & s \in \mathcal{X}_{\mathrm{u}} \\ +c, & s \in \mathcal{X}_+ \\ -c, & s \in \mathcal{X}_- \end{cases} \tag{5.87}$$

where

$$\nu = \frac{1}{X_{\mathrm{u}}}\left(F_{\mathrm{u}} + c(X_+ - X_-)\right) \tag{5.88}$$

$$F_{\mathrm{u}}(\rho) = \sum_{s\in\mathcal{X}_{\mathrm{u}}} p(s)\log(p(s)e^{\beta E_s}) \tag{5.89}$$

where $F_{\mathrm{u}}(\rho)$ is the (unnormalized) free energy calculated for the unbounded partition only. Note that $W^{(c)}(\rho)$ can be written in the more compact form

$$W^{(c)}(\rho) = \frac{1}{\beta}\log Z' - \frac{1}{\beta}\log\left(X_{\mathrm{u}}e^{-\beta\nu} + Z_+e^{\beta c} + Z_-e^{-\beta c}\right) \tag{5.90}$$

where

$$Z_\pm = \sum_{s\in\mathcal{X}_\pm} e^{-\beta E_s} \tag{5.91}$$

are the partition functions calculated over the positive and negative bounded partitions respectively. In Appendix 1.3 we find that in the optimal work extraction and state formation protocols the final / initial state of the system is the Gibbs state, which is diagonal in the energy eigenbasis. Hence equations (5.82) and (5.83) give the optimal work for arbitrary quantum states. For a two-level quantum system $d = 2$, the work content can be expressed succinctly as

$$W^{(c)}(\rho) = \tag{5.92}$$
$$\begin{cases} \frac{1}{\beta}\log Z' - \frac{1}{\beta}\log\left[e^{+\beta c}\left(1 + e^{-\beta(E+c/p(s_1))}\right)\right] & \text{if } c < \xi \\ \frac{1}{\beta}\log Z' - \frac{1}{\beta}\log\left[e^{-\beta c}\left(1 + e^{-\beta(E-c/p(s_1))}\right)\right] & \text{if } c > -\xi \\ \frac{1}{\beta}\log Z' + F(\rho) & \text{otherwise} \end{cases}$$

where

$$\xi = \frac{1}{\beta}\ln(1-p(s_1)) - F(\rho)\ . \tag{5.93}$$

Without loss of generality we have assumed above $p(s_1) \geq p(s_2)$ and we define $E = E_1 - E_2$. The above expression derives from (5.90) and the partitioning algorithm given in Appendix 1.2.

Returning to the c-bounded work of formation, (5.83) and the algorithm for finding the state space partition giving the c-bounded work of formation (5.83) is detailed in Appendix 1.2. Note that the Hamiltonian of S', the final state of the system, is given by $H'_{\rm S} = \sum_j E'_j \left|j\right\rangle\!\left\langle j\right|$. The work of formation for a two level system can be succinctly stated as

$$W_{\rm F}^{(c)}(\rho) = \begin{cases} \frac{1}{\beta}\log Z - F(\rho)\,, & c \geq \xi \\ \frac{1}{\beta}\log\left(p\, e^{\beta E'} Z\right) + c\frac{1-p}{p}\,, & c < \xi \end{cases} \tag{5.94}$$

where $\rho = p\left|0\right\rangle\!\left\langle 0\right| + (1-p)\left|1\right\rangle\!\left\langle 1\right|$, $p \geq 1/2$ and $H_{\rm S} = E\left|0\right\rangle\!\left\langle 0\right|$ is the Hamiltonian of the initial Gibbs state.

We now summarize some of the properties of the c-bounded work.

Corollary 3 (recovering the standard free energy)**.** The c-bounded work converages to the standard (Helmholtz) work in the limit $c \to \infty$.

$$\lim_{c\to\infty} W^{(c)}(\rho) = W^{(\infty)}(\rho) = -\Delta F$$

and

$$\lim_{c\to\infty} W_F^{(c)}(\rho) = W_F^{(\infty)}(\rho) = \Delta F$$

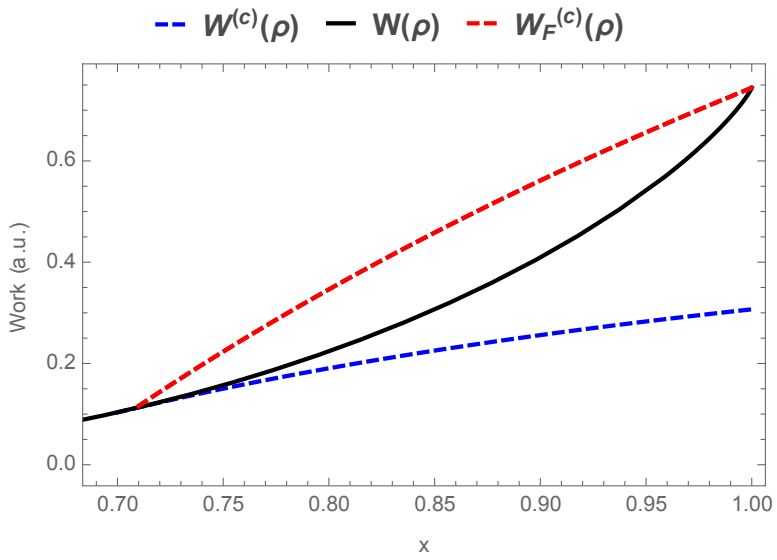


Figure 5.2: How c-bounded work extraction and state formation varies further from equilibrium. Figure shows the unbounded work $W(\rho)$, the c-bounded work content $W^{(c)}(\rho)$ and the c-bounded work of formation $W_F^{(c)}(\rho)$ for the state $\rho = p\,|0\rangle\langle 0| + (1-p)\,|1\rangle\langle 1|$ with Hamiltonian $H_S = E\,|0\rangle\langle 0|$ with $\beta = 1$, $E = 0.1$ and $c = 0.7$. In the above figure x denotes the probability $1-p$ of the system being in the excited state $|1\rangle$. Note that despite choosing a c-bound that allows for fluctuations of the order of the maximal work that can be extracted from the pure state $p = 1$, in general we can extract much less that this amount. There is a discontinuity in $W^{(c)}(\rho)$ at $x = 0$ where we recover $W^{(c)}(\rho) = W^{(\infty)}(\rho)$. Notice also that closer to the thermal state the dissipation (difference between the $W^{(c)}(\rho)$ or $W_F^{(c)}(\rho)$ and $W^{(\infty)}(\rho)$) is greater for state formations than work extraction, and this reverses as the state moves further from the Gibbs state.

Proof. Direct consequence of Lemma 5.62 □

Corollary 4 (recovering the single-shot free energy)**.** The c-bounded work converges to the single shot work in the limit $c \to 0$.

$$\lim_{c\to 0} W^{(c)}(\rho) = W_{\min}^{(0)}(\rho)$$

and

$$\lim_{c\to 0} W_F^{(c)}(\rho) = W_F^{\min}(\rho)$$

Proof. See Theorems 10 and 11 in Appendix .1.5 □

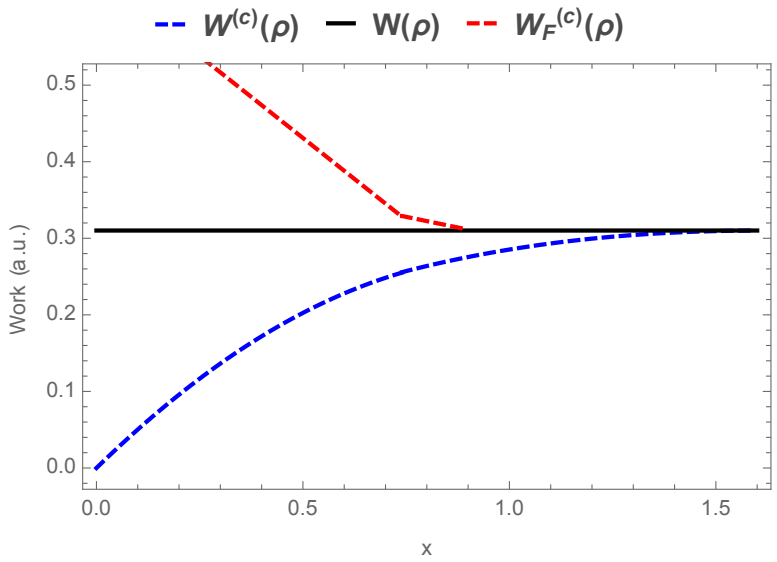


Figure 5.3: How c-bounded work extraction and state formation varies with c for a fixed initial / final state. figure shows $W_F^{(c)}(\rho)$ and $W^{(c)}(\rho)$ v.s. c for the trit state $\rho = 0.7\,|0\rangle\langle 0| + 0.2\,|1\rangle\langle 1| + 0.1\,|2\rangle\langle 2|$ with Hamiltonian $H_s = E_1\,|0\rangle\langle 0| + E_2\,|1\rangle\langle 1|$ with $E_1 = 0.1$, $E_2 = 0.2$. β is set to 1. x denotes the value of c, which we tune between 0 and 1.5. For small c the dissipation $|W^{(c)}(\rho) - W^{(\infty)}(\rho)|$ for the formation protocol is greater than for the extraction protocol, and for large c the relationship is inverted. Note that for $c > 0.9$ it possible to thermodynamically reversibly prepare state ρ but not to thermodynamically reversibly extract work from it.

Therefore the c-bounded work recovers the standard and single shot work in these limits for all initial and final states and Hamiltonia. We can use the c-bounded work to derive the following fundamental limitations on the magnitude and fluctuations of work that can be extracted from an infinite bath

Theorem 7. The c-bounded work is related to the non-fluctuating work by inequalities

$$W^{(c)}(\rho) \leq W^{(0)}(\rho) + c \tag{5.95}$$

$$W_{\mathrm{F}}^{(c)}(\rho) \geq W_{\mathrm{F}}^{(0)}(\rho) - c \tag{5.96}$$

becoming strict inequalities for $c > 0$

Proof: See lemmas 27 and 28 in Appendix .1.6.

These inequalities imply a fundamental trade-off between work and fluctuations. In order to do better than single shot work extraction / state formation, our work must have fluctuations that are greater than the increase in work / decrease in work cost with respect to the deterministic work. In Appendix 1.7 we show that the c-bounded work distributions that give $W^{(c)}(\rho)$ and $W_F^{(c)}(\rho)$ obey the Jarzinski equality [75]. In Supplementary Note 5 we prove that $\lim_{c\to 0} W^{(c)}(\rho) = W^{(0)}(\rho)$ and similarly for $W_{\mathrm{F}}^{(c)}(\rho)$.

It is simple to see that for any finite c a partial ordering of the states w.r.t work emerges. A simple way to observe this is to choose a qubit state ρ with Hamiltonian H_{S} and a thermal qubit state τ with Hamiltonian $H'_{\mathrm{S}} \neq H_{\mathrm{S}}$ such that neither state thermomajorizes the other (see [71] for examples). For any two such states there is a value of c below which $W^{(c)}(\rho)$ (for $\rho \to \tau$) is negative and $W_{\mathrm{F}}^{(c)}(\rho)$ (for $\tau \to \rho$) is positive, i.e. it costs work to perform both the forward and backwards transitions. Note that there are states with the same standard free energy that exhibit a partial order for finite c. Allowing the weight to fluctuate freely allows us to transition between these states freely. This is an example of how weight is not just a resource for extracting additional fluctuating work but in accommodating dynamics, even when on average its displacement remains zero. It is an interesting open question to determine how much we must allow work to fluctuate to allow a transition $\rho \to \sigma$ to be achieved without costing work.

Qubit Carnot engine. We now present an application of our result - computing the c-bounded Carnot efficiency for a single qubit Carnot engine model. This is the maximal efficiency the qubit engine can reach given that fluctuations in the work it produces are bounded by c. We use the same single qubit engine model as described in [124]. The engine operates by moving a qubit ρ with Hamiltonian $H_s = E\,|0\rangle\langle 0|$ between two baths of inverse temperature β_H and β_C, with $\beta_H < \beta_C$. The qubit has state $\rho_{H,C} = Z_{H,C}^{-1} e^{-\beta_{H,C}E}\,|0\rangle\langle 0| + Z_{H,C}^{-1}\,|1\rangle\langle 1|$ when in equilibrium with the hot / cold bath, where $Z_{H,C} = 1 + e^{-\beta_{H,C}E}$. The engine cycle begins with the qubit

in thermal equilibrium with the cold bath. In the first half of the cycle it is then placed in contact with the hot bath and work is extracted. In the second step of the cycle the qubit is returned to the cold bath and work is extracted a second time. In Appendix 2 we show that, in the case that fluctuations are not bounded, it is possible to reach Carnot efficiency with this engine, as shown in [124].

$$\eta_{\text{Carnot}} = 1 - \frac{\beta_H}{\beta_C} \tag{5.97}$$

In the case that c is finite, the work extracted in the first half of the cycle is given by

$$W_1^{(c)}(\rho) = \begin{cases} \frac{1}{\beta_H}\log\left(\frac{Z_H}{Z_C}\right) + \text{tr}[H_s\rho_C]\left(\frac{\beta_H-\beta_C}{\beta_H}\right), \text{ if } A > c \\ \frac{1}{\beta_H}\log\left(\frac{Z_H e^{\beta_H c}}{e^{c\beta_H}Z_C + e^{-E\beta_H}}\right), \text{ if } A \leq c \end{cases} \tag{5.98}$$

where

$$A = \frac{E}{Z_C}\left(\frac{\beta_C - \beta_H}{\beta_H}\right) \tag{5.99}$$

if $A \leq c$ we simply extract the difference in free energy between the two thermal states, otherwise we extract the c-bounded work of ρ_C in contact with the bath β_H. Similarly, on the second part of the cycle we extract

$$W_2^{(c)}(\rho) = \begin{cases} \frac{1}{\beta_C}\log\left(\frac{Z_C}{Z_H}\right) + \text{tr}[H_s\rho_H]\left(\frac{\beta_C-\beta_H}{\beta_H}\right), \text{ if } B > c \\ \frac{1}{\beta_C}\log\left(\frac{Z_C e^{-\beta_C c}}{e^{-c\beta_C}Z_H + e^{-E\beta_C}}\right), \text{ if } B \leq c \end{cases} \tag{5.100}$$

where

$$B = \frac{E}{Z_H}\left(\frac{\beta_C - \beta_H}{\beta_C}\right) \tag{5.101}$$

Note that satisfying (5.99) implies that (5.101) is also satisfied, therefore breaking inequality (5.101) is the condition for achieving Carnot efficiency in this model. Also note that B gives the minimum worst case fluctuation of the work extracted by this engine when operating thermodynamically reversibly. The efficiency is given by the ratio of the heat flow from the hot bath to the total work extracted in the cycle. The heat flow from the hot bath is found by applying the 1$^{\text{st}}$ law of thermodynamics,

$Q_H = \Delta\langle E\rangle(\rho_H \to \rho_C) + W_1^{(c)}$ where $\Delta\langle E\rangle(\rho_H \to \rho_C)$ is the change in the systems internal energy in the first part of the cycle. Therefore the c-bounded efficiency of the engine is given by

$$\eta^{(c)} = \frac{W_1^{(c)} + W_2^{(c)}}{\Delta U + W_1^{(c)}} \tag{5.102}$$

In the case that $c \to \infty$, we recover the Carnot efficiency, which is bounded from

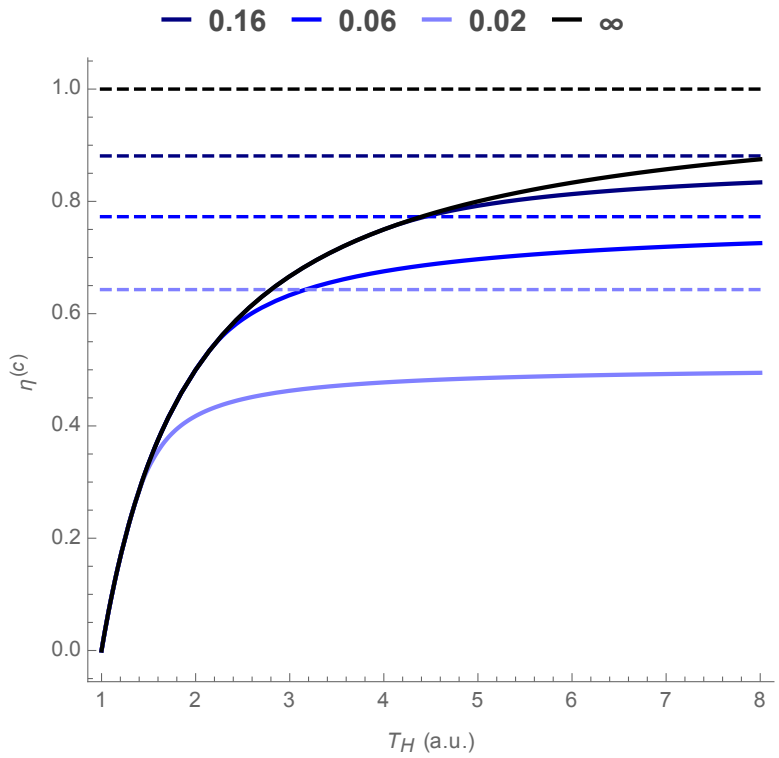


Figure 5.4: The corrected Carnot efficiency for various values of c for a single qubit heat engine. Figure shows the c-bounded Carnot efficiencies vs. T_H for single qubit engine with gap $E = 0.1$ and $T_C = 1$. The black line gives the unbounded Carnot efficiency. The Blue lines give the c-bounded efficiencies with there corresponding c's marked on the figure. The dashed lines give the maximum attainable efficiency in the limit of asymptotic temperature difference between $\beta_C/\beta_H \to \infty$

above by 1, i.e. we recover unit efficiency in the limit that $\beta_C/\beta_H \to \infty$. For any finite c this is no longer the case, with the maximal efficiency give by

$$\eta_{\max}^{(c)} = \lim_{\beta_C/\beta_H \to \infty} \eta^{(c)} = 1 - \frac{E}{2(2c+E)} \tag{5.103}$$

This gives an upper limit on the efficiency of the single qubit engine protocol described above, which is dependent only on the Hamiltonian of the qubit and the

parameter c.

As $E \to 0$, $\eta_{\max}^{(c)} \to 1$, but the work extracted tends to zero as the Gibbs states associated with the two bath temperatures become indistinguishable. For $c \to 0$ we get that $\eta_{\max}^{(c)}$ is bounded from below by $1/2$, but at $c = 0$ no work can be extracted as the thermal states are of full rank, giving $\eta_{\max}^{(0)} = 0$. Therefore we find that, although this engine cannot run at non-zero efficiency in the single-shot regime, if we allow for arbitrarily small fluctuations it is possible in principle to reach a maximum efficiency greater that $1/2$ (although the fluctuations in work will still be of the order of the work extracted, given inequality (5.95)). Similar results relating to the single-shot regime are discussed in [135].

In conclusion, we have derived tight bounds on the minimal fluctuations in work associated with thermodynamically reversible protocols, for which the average work is given by difference in free energy between initial and final states. We have shown in Theorem 5 that thermodynamically reversible protocols have fluctuations that diverge in size as the relative athermality of initial or final states increases. We then derived the c-bounded work content and work of formation of arbitrary quantum states, which can be understood as modified free energies that interpolate between the standard and single shot free energies. We have seen that, for any finite c, some of the phenomenology of single-shot thermodynamics (namely thermodynamic irreversibility and a partial order of states with respect to work) are present for some transitions. Furthermore we have found that it is impossible to extract more than the deterministic work content of a system without necessitating fluctuations that are greater than the gain in work (and similarly for the work cost of state formation).

An interesting open question is to what extent we must allow work to fluctuate in order to allow for a given state transformation. Answering this question would require the extension of these results to processes with arbitrary initial and final states, including the case where both initial and final states contain coherences between energy levels, which is ultimately beyond the scope of this thesis. We have

also used the c-bounded work to study how bounding work fluctuations affects the efficiency of a single qubit heat engine, and have derived an upper bound on efficiency of this engine that depends only on c and the engine's Hamiltonian, establishing a fundamental trade-off between a the engines efficiency and the fluctuations in the work it produces. This is particularly pertinent when considering any realistic small-scale thermal machine, which will only be able to perform thermodynamic protocols that result in small fluctuations in the energy of the work battery.

All the the results derived thus far assume that the system can interact with all the microstates of an infinite thermal bath. We now turn our attention to the final chapter of this thesis, where we remove this strong assumption and re-derive the second law in the case of a finite thermal bath.

5.4 Finite bath corrections to the second law

In this section we redress our assumption that the thermal bath (Definition 24) has infinite volume. The results presented in this section have been published as [113]. We consider the limit wherein the bath is a very large but finite system. In fact, even if the bath is still technically infinite in volume, we should still treat it as a large finite system. This is because as the system interacts with the bath over a finite period of time, it will only be able to probe a finite volume of the bath[6]. We now restate our assumptions on the heat bath, and proceed as before without taking the infinite limit.

Definition 24: We define a heat bath as a large system with a continuous energy spectrum defined by a Hamiltonian H_B. The density of microstates of the bath is given by $\Omega(E, V) = e^{S(E,V)}$ where $S(E, V)$ is the entropy of the microcanonical

[6]For example, the interactions between the system and bath are ultimately bounded to lie within the light cone of the system, the speed of information transfer is bounded by the Lieb-Robinson bounds [87] and, more practically, heat transfer rate is limited by the thermal conductivity of the bath

ensemble for a given energy E and (dimensionless) volume V. Furthermore, we demand that

1. The entropy is extensive, $S(kE, kV) = kS(E, V)\ \forall\ k > 0$
2. The dimensionless volume V is large
3. The bath is in a Gibbs state with inverse temperature β, s.t. a microstate with energy E has an associated probability $Z_B^{-1} e^{-\beta E}$

Assumption 1 (extensivity) implies that the microcanonical entropy varies linearly with volume,

$$S(E, V) = V f(u), \quad u = E/V \tag{5.104}$$

where $f(u)$ is some function of the energy density. Using $\Omega(E) \propto e^{S(E)}$, and using the probability of microstates $p(x) \propto e^{-\beta H(x)}$ (equation (5.4), we assume the bath microstates are still well described by a Gibbs distribution), the distribution of the bath energy densities obeys

$$p(u) \propto e^{V(f(u) - \beta u)} \tag{5.105}$$

In the large V limit we can apply the saddle point approximation to approximate $p(u)$ by expanding the exponent $g(u) = f(u) - \beta u$ about its maxima (hence discarding the linear expansion term), $g(u) \sim g(u_\beta) + 1/2(u - u_\beta)^2 g''(u_\beta)$, where u_β is the absolute maxima of $g(u)$ as a function of u. This implies that $f'(u_\beta) = \beta$ (to achieve $g'(u) = 0$) and $f''(u_\beta) < 0$.

$$\begin{aligned} p(u) &\propto e^{V\left(f(u_\beta) + \frac{1}{2} f''(u_\beta)(u - u_\beta)^2\right)} \\ &\propto e^{-\frac{V}{2}\left|f''(u_\beta)\right|(u - u_\beta)^2} \end{aligned} \tag{5.106}$$

This is a Gaussian with mean $\langle u \rangle = u_\beta$ and variance $\langle (u - u_\beta)^2 \rangle = |V f''(u_\beta)|^{-1}$. We can relate $f''(u_\beta)$ to the heat capacity of the bath by

$$C := V \frac{\partial \langle u \rangle}{\partial T} = -\frac{V}{T^2} \frac{\partial \langle u \rangle}{\partial \beta} \tag{5.107}$$

Starting with $f'(u_\beta) = \beta$ and differentiating w.r.t β gives $f''(u_\beta) = 1/\left(\frac{\partial u_\beta}{\partial \beta}\right)$. substituting this into (5.107) yields

$$C = -\frac{V}{T^2} \frac{1}{f''(u_\beta)} \tag{5.108}$$

Note that $f''(u_\beta) < 0$ implies that the heat capacity is positive, as is always the case in "ordinary matter"[7]. Also note that $C \propto V$, because $f(u_\beta)$ is independent of V. The variance of $p(u)$ can now be expressed as

$$\left\langle (u - \langle u \rangle)^2 \right\rangle^{1/2} = \frac{T\sqrt{C}}{V} \propto V^{-1/2} \ , \tag{5.109}$$

which is small when V is large. Finally, we use our approximate form for $p(u)$ to approximate the density of states of the bath, $\Omega(E, V) \propto e^{\beta E} p(u)$, as

$$\Omega(E, V) \propto \exp\left(\beta E - \frac{\gamma E^2}{2}\right), \tag{5.110}$$

where we have rescaled the energy such that $\langle E \rangle = 0$, and we define $\gamma = \frac{1}{CT^2}$.

5.4.1 Finite bath corrections to the second law

In this section we analyze the transformation power of thermal operations in the case of a finite bath as outlined in the previous section. First we recap Lemma 20, which provides a correction to the second law in terms of the mutual information

[7] A counterexample to this is semi-classical black holes, which have a negative heat capacity [93]

between system and bath following the thermal operation

$$W_{\rho\to\sigma} = F(\rho) - F(\sigma) - \frac{1}{\beta}\mathcal{I}(\mathcal{E}(\rho_S \otimes \rho_B)) \tag{5.111}$$

Note that in our derivation of Lemma 20 we make no assumption on the density of states of the bath, other than it being convex. This necessary and sufficient condition is not always useful, because it involves the final state of the bath. However, it precisely articulates the effect of having a finite bath. Note that a similar bound was previously found in [50] and [111], albeit under more significant assumptions such as the bath being finite dimensional and having certain Hamiltonian properties.

In the previous section we showed that when the heat capacity of the bath is infinite, it is possible to extract average work equal to the free energy difference. In this case, the system-bath final state is of the form $P(s', E') = P(s')p_b(E')$, i.e. the system and bath end up uncorrelated, and the bath remains in a thermal state. The mutual information correction of Lemma 20 quantifies the correlations in the final joint system-bath state $P(s', E')$. We now show that this idealized situation cannot be achieved with a finite bath, and the corrections to the second law that we derive can be thought of as quantifying the minimal correlations that can be achieves in the final system-bath state.

The bound in (5.111) requires detailed knowledge of the final system and bath joint state which is not typically available in a realistic setting (in order to compute the mutual information). This isn't an ideal situation, as this detailed knowledge is not typically available, and we are interested in calculating the optimal work in the case of a large but finite bath. Equation (5.111) lets us interpret the bounds we derive as the minimal mutual information generate between system and finite bath in a given state transition. We now present a necessary condition which only depends on the state of the system and the heat capacity of the bath C.

Theorem 8 (A necessary condition for the possibility of a transformation is)**.**

$$\langle W \rangle \leq -\Delta F - \frac{1}{2\beta C}\Delta S^2 \ , \tag{5.112}$$

where ΔS is the change in entropy of the system.

Proof. See Appendix .3.1.

This bound is simple to calculate, and the correction to the standard second law $(2\beta C)^{-1}\Delta S^2$ is quadratic in the entropy change between initial and final states, giving a simple measure of the thermodynamic irreversibility inherent in all state transformations that change the entropy of the system. From this, we can conclude that

Thermodynamic reversibility requires an infinite heat bath.

This upper bound to the extractable work may not be tight in general, although we explore some cases for which it is. We present tight upper bounds that only depends on the initial and final states of the system, like (5.112), for the case that the initial and/or final state is maximally mixed, $p(s) = \mathsf{const}$ or $q(s') = \mathsf{const}$. When the initial state is maximally mixed, this corresponds precisely to the often studied case of Landauer erasure [84, 20]. For these cases we can refine the work upper bound in (5.112) to a tight upper bound and show that it is achievable through operations that are independent of the bath energy (See Appendix B3).

Theorem 3: *When the final state of the system is maximally mixed, the necessary and sufficient condition is*

$$\langle W \rangle \leq -\Delta F - \frac{1}{2\beta C}\Delta S^2 \ , \tag{5.113}$$

up to first order in $1/C$*. When the initial state is maximally mixed, the necessary and sufficient condition is*

$$\langle W \rangle \leq -\Delta F - \tfrac{1}{2\beta C}\left(\Delta S^2 + \mathsf{Var}[\rho]\right) \ , \quad (5.114)$$

up to first order in $1/C$*, where* ρ *is the final state of the system. The 'varentropy' is defined as*

$$\mathsf{Var}[\rho] = \sum_s p(s) \log^2 p(s) - \left[\sum_s p(s) \log p(s)\right]^2 , \quad (5.115)$$

and is always positive.

Proof. See Appendix B3.

Equation (5.114) gives a tight upper bound to the work cost of Landauer erasure in the case of a large but finite bath. Note that the correction contains a *varentropy* term. Varentropy is best understood by considering the entropy as the expectation value of the 'surprise' (also known as the self-information, or 'surprisal') function in Shannon information theory, $S(\rho) = \langle -\log p(s) \rangle$. The surprise is the information gain when sampling state s from a given prior distribution $p(s)$. By (5.115), the varentropy is simply the variance of the surprise. This expression has turned up in non-asymptotic Shannon theory (for example [82, 77, 104]), but this is (to my knowledge) its first appearance in the context of the second law. As the second law for an infinite bath is phrased in terms of the entropy (a first-order information measure), its is perhaps unsurprising that non-asymptotic information measures are required when we remove the asymptotic approximation of an infinite bath.

For thermal states $p(s) \propto e^{-\beta \mathcal{E}_s}$, the varentropy is precisely the heat capacity of the system (5.107) of the system (see section 2.2.2 of [112]). For example, consider the case where we prepare a thermal state of an arbitrary Hamiltonian H'_S from the thermal state of a trivial Hamiltonian. The term $\mathsf{Var}[p(s')]$ gives the heat capacity of

the prepared system and the correction includes the ratio of this to the heat capacity of the bath.

$$\langle W \rangle \leq -\Delta F - \frac{1}{2\beta C_B}\Delta S^2 + \frac{1}{2\beta}\frac{C_S}{C_B} \tag{5.116}$$

5.5 Conclusions

We have derived upper bounds for the optimal work associated with thermal operations where the bath is large but finite, and tight upper bounds in the case of Landauer erasure and its reverse process. Our approach has been general in the sense that we do not need to consider the particular microscopic structure of the heat bath (e.g. whether it is made of fermions, bosons,...). The only quantities that play a role are its temperature and its heat capacity. When the heat capacity diverges, i.e. the bath becomes infinite, one recovers all the standard results, such as the result in [124, 2].

Other work on finite-size limitations includes [111, 72, 107, 135]. For example in [111] tight corrections to the Landauer bound are found in terms of the dimension of the bath, recovering the Landauer limit $\langle W \rangle \leq \Delta S$ when the heat bath is infinite dimensional. There is however a clear difference between bounding the dimensionality of the Hilbert space of the bath, and bounding its volume as we do. To illustrate this difference, there are situations where the bath is infinite dimensional but with a finite volume and heat capacity (for example a bosonic bath). In these cases bounding dimension results in trivial corrections to the free energy whereas bounding the heat capacity provides non-trivial corrections. Furthermore, the fact that the bath has infinite dimensions is also a necessary condition for the appearance of an infinite recurrence time, which as previously discussed is needed for the emergence of irreversibility [21] [8]. Working in the regime of finite but large environment, we find

[8] A simple argument goes, if an isolated system *must* return ϵ-close to its initial state, then

corrections to the second law that are universal, in that they apply to all conceivable environments, regardless of their constituents, Hamiltonian and Hilbert space dimension.

One possible area where these results could be applied is to exploring the behaviour of heat engines operating in finite time. We have observed that the effect of having access to a finite bath has a marked effect on the minimal achievable dissipation in some protocols, suggesting that there could be significant finite size effects in small scale thermal engines. A correction to the free energy should also correspond to a correction of the Carnot efficiency of ideal cyclic process. Previous work such as [135, 123, 131] has looked at this question for particular models. An open question is to determine the work-optimal processes for arbitrary state transformations with a finite bath. Whereas for an infinite bath the work is a function of the state, and as a result the work is "path independent", for finite baths our results suggest the existence of unique optimal processes. For example, a Szilard engine [129] operates cyclically on systems with trivial Hamiltonian, and so the optimal work is captured exactly by equations (5.113) and (5.114). In the case when the dimension of the working system is greater than two, the variantropy term in (5.114) differs for states of equal entropy, and so there will be optimal paths through the state space of the system which minimize varentropy - a constraint that isn't present in the infinite bath analysis due to path independence. Determining these optimal paths it tantamount to determining the optimal heat engine protocols operating in finite time /space, which is an open problem of practical importance.

any transition from that state cannot be seen as irreversible, as one can return without expending energy

Appendices

.1 Thermal operations

In this Appendix we derive the c-bounded work and related results.

.1.1 determining the optimal map

Lemma 23. We can always find an optimal protocol where $t(s'|s) \to \tilde{t}_s = e^{-\beta E_{s'}}/Z'$, where $Z' = \mathrm{tr}[e^{-\beta H'_S}]$, and $w(s',s) \to \tilde{w}_s$ that obeys $|\tilde{w}(s) - \langle W\rangle| \leq |w(s',s) - \langle W\rangle|$ $\forall\ s, s'$

Proof. The average work extracted by a given protocol is given by

$$W = \sum_{ss'} p(s)t(s'|s)w(s',s) \tag{117}$$

where $\{t(s'|s), w(s',s)\}$ obey constraints (5.70) (5.71) and (5.73).

Assume there exists some optimal protocol $\{t(s'|s), w(s',s)\}$ that satisfies these and obeys the c-bound

$$|w(s',s) - W| \leq c \quad \forall s,\ s' \tag{118}$$

Define a new protocol with

$$\tilde{t}(s') = \frac{e^{-\beta E_{s'}}}{Z'} \tag{119}$$

$$\tilde{w}(s) = \sum_k t(k|s)w(k|s) \tag{120}$$

The average work extracted by this protocol is

$$\tilde{W} = \sum_{ss'} p(s)\tilde{t}(s')\tilde{w}(s) = \sum_s p(s)\tilde{w}(s) = \sum_{ss'} p(s)t(s'|s)w(s',s) \tag{121}$$

therefore it extract the same amount of work as the optimal protocol. It is also

obeys the microscopic reversibility inequality (5.73) as

$$\begin{aligned}\sum_s \tilde{t}(s')e^{\beta(\tilde{w}(s)-\Delta E_{ss'})} &= \frac{1}{Z'}\sum_s e^{\beta\left(\sum_k t(k|s)w(k|s)-E_s\right)} \\ &\leq \frac{1}{Z'}\sum_{sk} t(k|s)e^{\beta w(k|s)-\beta E_s} \\ &= \frac{1}{Z'}\sum_{sk} t(k|s)e^{\beta w(k|s)-\beta E_s+\beta E_k-\beta E_k} \\ &\leq \frac{1}{Z'}\sum_k e^{-\beta E_k} \\ &= 1\end{aligned}$$

where we have used the fact that $\sum_k t(k|s) = 1$ and the convexity of the exponential function to get $e^{\beta\tilde{w}(s)} \leq \sum_k t(k|s)e^{\beta w(k|s)}$, and in the second to last line have used the reversibility of the original map. Finally, the new protocol also obeys the c-bound as the work values $\tilde{w}(s)$ are convex sums of the work values $w(s', s)$ and therefore $\max\{w(k|s)\} \geq \tilde{w}(s) \geq \min\{w(k|s)\}$. As W is unchanged then the worst case fluctuations of the new work distribution about the average must be less than or equal to those of the optimal distribution. □

We now drop the tilde from $\tilde{w}(s)$ and $\tilde{t}(s')$. It is simple to check that $t(s') = 1/Z'e^{-\beta E_{s'}}$ satisfies the conditions (5.70) and (5.71). The reversibility inequality (5.73) becomes

$$Z' \geq \sum_s e^{\beta(w(s)-E_s)} \tag{122}$$

and the average work is given by

$$W = \sum_s p(s)w(s) \tag{123}$$

We are now in a position to derive $W^{(c)}(\rho)$. The Lagrangian is the same as employed in the previous section except that it includes terms bounding all $|w(s) -$

$\sum_k p(s_k)w(k)| \leq 0$. Due to the exponential in the reversibility term $Z' \geq \sum_s e^{\beta(w(s)-E_s)}$ we must linearise these bounds so as to avoid generating a transcendental equation when extremizing the Lagrangian over $w(s)$. We therefore select the bounds

$$w(s) - \sum_k p(s_k)w(k) \leq c \tag{124}$$

$$\sum_k p(s_k)w(k) - w(s) \leq c \tag{125}$$

with associate Lagrange parameters $\mu_{ss'}$ and $\bar{\mu}_{ss'}$. As all fluctuations are either positive ($w(s) \geq W^{(c)}(\rho)$) or negative ($w(s) < W^{(c)}(\rho)$) one of these bounds will always be trivial for each work value $w(s)$. Including the reversibility constraint (122) we maximise the average work $W = \sum_s p(s)w(s)$ by extremizing the Lagrangian

$$\mathcal{L} = \sum_s f_s w(s) + \lambda \left(Z' - \sum_s e^{\beta(w(s)-E_s)} \right) + c \sum_s (\mu_s + \bar{\mu}_s) \tag{126}$$

where

$$f_s = p(s)\, f - (\bar{\mu}_s - \mu_s) \tag{127}$$

where $f = 1 + \sum_k (\bar{\mu}_k - \mu_k)$ and we have replaced the old Lagrange parameters with $\lambda = Z' \sum_s \lambda_s$ and $\mu_s(\bar{\mu}_s) = \sum_{s'} \mu_{ss'}(\bar{\mu}_{ss'})$. Maximizing w.r.t $w(i)$ and λ gives

$$w(s) = \frac{1}{\beta} \log \left(\frac{f_s e^{\beta E_s}}{\beta \lambda} \right) \tag{128}$$

$$\lambda = \frac{1}{\beta Z'} \tag{129}$$

where we have used $\sum_s f_s = 1$. Substituting these into the Lagrangian simplifies it to

$$\mathcal{L} = \frac{1}{\beta} \sum_s f_s \log \left(f_s e^{\beta E_s} Z' \right) + c \sum_s (\mu_s + \bar{\mu}_s) \tag{130}$$

Maximizing w.r.t μ_j and $\bar{\mu}_j$ under the condition $\mu_j \geq 0$ and $\bar{\mu}_j \geq 0$ gives

$$\frac{\partial \mathcal{L}}{\partial \mu_j} = 0 \quad \rightarrow \quad \log\left(f_j e^{\beta E_j}\right) - \sum_s p(s) \log\left(f_s e^{\beta E_s}\right) = -\beta c \tag{131}$$

$$\frac{\partial \mathcal{L}}{\partial \bar{\mu}_j} = 0 \quad \rightarrow \quad -\log\left(f_j e^{\beta E_j}\right) + \sum_s p(s) \log\left(f_s e^{\beta E_s}\right) = -\beta c \tag{132}$$

where we have used

$$\frac{\partial f_s}{\partial \mu_j} = -p(s) + \delta_{sj} \tag{133}$$

$$\frac{\partial f_s}{\partial \bar{\mu}_j} = p(s) - \delta_{sj} \tag{134}$$

For $c > 0$, equations (131) and (132) cannot be simultaneously satisfied by any $w(j) = 1/\beta \log\left(f_j e^{\beta E_j} Z'\right)$ therefore if $\mu_j > 0$ then $\bar{\mu}_j = 0$ and visa-versa. As $w(i) = 1/\beta \log\left(f_i e^{\beta E_i} Z'\right)$ equations (131) and (132) can be written in the form

$$\mu_j \neq 0 \rightarrow w(j) - \langle W \rangle = -c \tag{135}$$

$$\bar{\mu}_j \neq 0 \rightarrow w(j) - \langle W \rangle = c \tag{136}$$

Therefore fluctuations saturate either a positive or negative bound, or saturate no bounds ($\mu_j = \bar{\mu}_j = 0$). It will therefore be useful to partition the set of energy levels into those for which the resulting fluctuations will saturate a positive bound $i \in \mathcal{X}_+$, a negative bound $j \in \mathcal{X}_-$ or saturate no bounds $u \in \mathcal{X}_u$. For work values $w(i)$, $w(i')$ that saturate a positive bound and $w(j)$, $w(j')$ that saturate a negative bound we

have

$$
\begin{aligned}
f_i e^{\beta E_i} &= f_{i'} e^{\beta E_{i'}} && (137)\\
f_j e^{\beta E_j} &= f_{j'} e^{\beta E_{j'}} && (138)\\
f_j e^{\beta E_j} &= f_{i'} e^{\beta (E_{i'} - 2c)} && (139)\\
f_i &= p(s_i) f - \bar{\mu}_i && (140)\\
f_j &= p(s_j) f + \mu_j && (141)\\
f_u &= p(s_u) f && (142)
\end{aligned}
$$

where (137) comes from summing (131) for μ_j, $\mu_{j'}$, (138) from summing (132) for $\bar{\mu}_i$, $\bar{\mu}_{i'}$, (139) comes from subtracting (131) for μ_j and (132) for $\bar{\mu}_i$ and (140)–(142) are just the definition (127) with the conditions $\bar{\mu}_j = 0$, $\mu_i = 0$ and $\mu_u = \bar{\mu}_u = 0$ applied. (137)–(139) and (142) allow us to simplify (131) and (132) to

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mu_j} &= 0 \rightarrow \log\left(\frac{f}{f_j e^{\beta E_j}}\right) = \nu + c && (143)\\
\frac{\partial \mathcal{L}}{\partial \bar{\mu}_i} &= 0 \rightarrow \log\left(\frac{f}{f_i e^{\beta E_i}}\right) = \nu - c && (144)\\
\nu &= \frac{\beta}{X_u}\left(\frac{1}{\beta} H_u(\rho) + c(X_- - X_+) - \langle E_u(\rho) \rangle\right) && (145)
\end{aligned}
$$

Where $X_+ = \sum_{s \in \mathcal{X}_+} p(s)$, $X_- = \sum_{s \in \mathcal{X}_-} p(s)$, $X_u = \sum_{s \in \mathcal{X}_u} p(s)$, $H_u(\rho) = -\sum_{s \in \mathcal{X}_u} p(s) \log p(s)$ and $\langle E_u(\rho) \rangle = \sum_{s \in \mathcal{X}_u} p(s) E_s$. (137)–(142) let us to relate the remaining Lagrange parameters by

$$
\begin{aligned}
\bar{\mu}_{i'} &= f(p(s_{i'}) - p(s_i) e^{\beta(E_i - E_{i'})}) + \bar{\mu}_i e^{\beta(E_i - E_{i'})} && (146)\\
\mu_{j'} &= f(p(s_j) e^{\beta(E_j - E_{j'})} - p(s_{j'})) + \mu_j e^{\beta(E_j - E_{j'})} && (147)\\
& && (148)\\
\bar{\mu}_i &= \frac{(k - \mu_j)(e^{\beta E_i} p(s_i) - e^{\beta(2c - E_j)} p(s_j)) - \mu_j e^{\beta(2c - E_j)}}{(1 - p(s_i)) e^{\beta E_i} + p(s_j) e^{\beta(2c - E_j)}}
\end{aligned}
$$

where $k = f - \bar{\mu}_i + \mu_j$. Summing (146) over i' and (147) over j' gives

$$\sum_{i'} \bar{\mu}_{i'} = f(X_+ - p(s_i)e^{\beta E_i} Z_+) + \bar{\mu}_i e^{\beta E_i} Z_+ \tag{149}$$

$$\sum_{j'} \mu_{j'} = f(p(s_j)e^{\beta E_j} Z_- - X_i) + \mu_j e^{\beta E_j} Z_- \tag{150}$$

where $Z_+ = \sum_{s \in \mathcal{X}_+} e^{-\beta E_s}$ and $Z_- = \sum_{s \in \mathcal{X}_+} e^{-\beta E_s}$. $f = 1 + \sum_{i'} \bar{\mu}_{i'} - \sum_{j'} \mu_{j'}$ therefore we can get f in terms of $\bar{\mu}_i$ and μ_j alone

$$f = \frac{1 + \bar{\mu}_i e^{\beta E_i} Z_+ - \mu_j e^{\beta E_j} Z_-}{p(s_i)e^{\beta E_i} Z_+ + p(s_j)e^{\beta E_j} Z_- + X_u} \tag{151}$$

We now have f in terms of $\bar{\mu}_i$ and μ_j, and (148) relates these to eachother, so we can solve (143) and (144) simultaneously (using $k = f - \bar{\mu}_i + \mu_j$) to find $\bar{\mu}_i$ and μ_j, and by (146), (147) all Lagrange multipliers. Substituting (146) and (147) into (137) and solving for μ_j gives

$$\mu_j = \frac{k e^{\beta E_i}(e^{-\beta c} - p(s_j)e^{\beta E_j + \nu})}{e^{\beta(E_i - c)} + e^{\beta(E_j + c)} + (1 - p(s_i) - p(s_j))e^{\beta(E_i + E_j) + \nu}} \tag{152}$$

and similarly for $\bar{\mu}_i$

$$\bar{\mu}_i = \frac{k e^{\beta E_j}(e^{\beta c} - p(s_i)e^{\beta E_i + \nu})}{e^{\beta(E_i - c)} + e^{\beta(E_j + c)} + (1 - p(s_i) - p(s_j))e^{\beta(E_i + E_j) + \nu}} \tag{153}$$

Using (151) and $k = f - \mu_j + \bar{\mu}_i$ simultaneously solve (152) and (153) to give

$$\bar{\mu}_i = \frac{p(s_i)e^{\nu} - e^{-\beta(E_i - c)}}{e^{\nu} X_u + Z_+ e^{\beta c} + Z_- e^{-\beta c}} \tag{154}$$

$$\mu_j = \frac{e^{-\beta(E_j + c)} - p(s_j)e^{\nu}}{e^{\nu} X_u + Z_+ e^{\beta c} + Z_- e^{-\beta c}} \tag{155}$$

it is simple to check that (146) and 147 result in exactly the same equation for all $\bar{\mu}_{i'}$ and $\mu_{j'}$ but with the corresponding index. Now armed with the explicit form of the Lagrange parameters we have solved the Lagrangian. It is easy to check that for these solutions for $\bar{\mu}_i$, μ_j the Lagrangian simplifies to $\mathcal{L} = \sum_s p(s)w(s)$ where

$w(s) = 1/\beta \log(f_s e^{\beta E_s} Z')$ where the f_i are now of the form

$$f_s = \gamma^{-1} \begin{cases} p(s)e^{\nu}\,, & s \in \mathcal{X}_u \\ e^{\beta(c-E_s)}\,, & s \in \mathcal{X}_+ \\ e^{-\beta(c+E_s)}\,, & s \in \mathcal{X}_- \end{cases} \tag{156}$$

where

$$\gamma = e^{\nu} X_u + Z_+ e^{\beta c} + Z_- e^{-\beta c} \tag{157}$$

Substituting in the above values for f_s into $\mathcal{L} = 1/\beta \sum_s p(s) \log(f_s e^{\beta E_s Z'})$ gives our final result for work extraction

$$W^{(c)}(\rho) = \frac{1}{\beta} \left(\log Z' - \log \gamma \right) \tag{158}$$

Next we show that γ can be written as $\gamma = \sum_s p(s)^0 e^{-\beta E_s} e^{\beta \theta_s}$ there θ_S is the fluctuation associates with subspace $|s\rangle\langle s|$ which, in the c-bounded distribution, is given by

$$\theta_s = \begin{cases} \frac{1}{\beta} \log(p(s)e^{\beta E_s}) + \frac{\nu}{\beta}, & s \in \mathcal{X}_u \\ +c, & s \in \mathcal{X}_+ \\ -c, & s \in \mathcal{X}_- \end{cases} \tag{159}$$

In general, the c-bounded work is given by the difference between the free energy of the final thermal state and the c-bounded free energy

$$F^{(c)}(\rho) = -\log \sum_s p(s)^0 e^{-\beta E_s} e^{\beta \theta_s} \tag{160}$$

.1.2 Finding the optimal partition

In this appendix we derive the inequalities for partitioning the state space into positively bounded $\mathcal{X}_+$, negatively bounded $\mathcal{X}_-$ and unbounded $\mathcal{X}_u$ energy levels, as required by our main result. We derive the general set of inequalities for any state

and Hamiltonian and give worked through examples of how to find the partition for arbitrary 2 and 3 dimensional systems. Firstly we derive the partition inequalities for work extraction protocols, and then show that the partition inequalities for work of formation are identical.

The Lagrangian is maximised under the condition that $\bar{\mu}_i$ and μ_j given in (152) and (153) are positive. Therefore if our optimization gives a negative Lagrange parameter it is set to zero, removing the corresponding c-bound. The Lagrange parameters are positive when the following inequalities are satisfied

$$\frac{1}{\beta}\log\left(p(s_i)e^{\beta E_i}\right) > c - \frac{\nu}{\beta} \quad \rightarrow \bar{\mu}_i > 0 \tag{161}$$

$$\frac{1}{\beta}\log\left(p(s_j)e^{\beta E_j}\right) < -c - \frac{\nu}{\beta} \quad \rightarrow \mu_j > 0 \tag{162}$$

where

$$\nu = \frac{\beta}{X_u}\left(\frac{1}{\beta}H_u(\rho) + c(X_- - X_+) - \langle E_u(\rho)\rangle\right) \tag{163}$$

and we have used $\gamma \geq 0$. Clearly ν depends on how you partition the state space in to positively, negatively and unbounded fluctuations. In the following we derive a set of inequalities that determine the unique partition give ρ and c. The following observations simplify the problem

Lemma 24. For any c-bounded work distribution where bounded fluctuations saturate their bounds, $X_\pm < 1/2$

Proof. Consider a c-bounded work distribution with work values $\{w(s)\}$, average work $W = \sum_s p(s)w(s)$, and all work values obey $|w(s) - W| \leq c$. Break the work distribution up into work values that give positive fluctuations $w(i) \geq W$ and negative fluctuations $w(j) < W$. Writing the fluctuations as $\theta_i = w(i) - W$ and

$\theta_j = W - w(j)$ and substituting into $W = \sum_s p(s)w(s)$ we get

$$\sum_i p(s_i)\theta_i = \sum_j p(s_j)\theta_j \tag{164}$$

which merely states that the average (non-absolute value) fluctuation of a random variable is zero, as is always the case. Take the case that $\sum_{s\in\mathcal{X}_+} p(s) > \sum_{s\in\mathcal{X}_-} p(s)$ and $X_+ > 1/2$. Clearly $\sum_{s\in\mathcal{X}_-} p(s) < 1/2$ so $X_- < 1/2$. If $X_+ \geq 1/2$ then, as bounded fluctuations saturate their bounds, (164) gives the inequality

$$\sum_{s\in\mathcal{X}_-} p(s)c < \sum_{s\in\mathcal{X}_+} p(s)\theta_s \tag{165}$$

taking the factor of $\sum_{s\in\mathcal{X}_+} p(s)$ to the other side we can write the inequality

$$c < \frac{1}{\sum\limits_{s\in\mathcal{X}_-} p(s)} \sum_{s\in\mathcal{X}_-} p(s)\theta_s \tag{166}$$

The right hand side is a convex sum, and all $\theta_s \geq 0$, therefore at least one $\theta_s > c$ contradicting the fact that the work distribution is c-bounded. Therefore we must have $X_+ < 1/2$. A similar argument for the case $\sum_{s\in\mathcal{X}_+} p(s) < \sum_{s\in\mathcal{X}_-} p(s)$ gives that $X_- < 1/2$. □

The second observation is that the inequalities (161) and (162) obey a β-ordering hierarchy. The β-odered state, $\rho^{\downarrow\beta}$, is defined as

$$\rho^{\downarrow\beta} = (p(s_1), p(s_2), \ldots, p(s_R)) \quad , \quad p(s)e^{\beta E_s} \geq p(s_{s+1})e^{\beta p(s_{s+1})} \tag{167}$$

where $R = \text{rank}(\rho)$ (there is no work value $w_s^{(\infty)}$ associated with $p(s) = 0$). The β-ordering gives $w_i^{(\infty)} \geq w_{i+1}^{(\infty)}$. Therefore if $w_i^{(\infty)}$ satisfies (161) then so does $w_{i-1}^{(\infty)}$, and if $w_j^{(\infty)}$ satisfies (162) the so does $w_{j+1}^{(\infty)}$. From this we can deduce that the

partition of the state space will look like

$$\rho^{\downarrow\beta} = (\underbrace{p(s_1), p(s_2), \ldots, p(s_k)}_{\mathcal{X}_+}, \underbrace{p(s_{k+1}), \ldots, p(s_{l-1})}_{\mathcal{X}_u}, \underbrace{p(s_l), \ldots, p(s_R)}_{\mathcal{X}_-}) \tag{168}$$

Finally, it will be useful to put the inequalities in to the following form. In the following we drop the (∞) superscript from the unbounded work values $w_s^{(\infty)}$. Subscript i will label positively bounded fluctuations and subscript j negatively bounded fluctuations. (161) and (162) can be written in terms of the unbounded work distribution only (therefore the problem of finding the partition, assuming it exists, is a closed form)

$$\begin{aligned} W^{(\infty)}(\rho) < X_u(w(i') - c) + \sum_{s\in\mathcal{X}_+} p(s)(w(s) - c) \\ + \sum_{s\in\mathcal{X}_-} p(s)(w(s) + c) \quad \forall\, i' \in \mathcal{X}_+ \end{aligned} \tag{169}$$

$$\begin{aligned} W^{(\infty)}(\rho) > X_u(w(j') + c) + \sum_{s\in\mathcal{X}_+} p(s)(w(s) - c) + \sum_{s\in\mathcal{X}_-} p(s)(w(s) + c) \\ \forall\, j' \in \mathcal{X}_- \end{aligned} \tag{170}$$

where $W^{(\infty)}(\rho)$ is the unbounded average work $W^{(\infty)}(\rho) = \beta^{-1}\log Z' + F(\rho)$ and $X_u = 1 - X_+ - X_-$. We now prove that for a given ρ and c there exists a unique partition for which all inequalities (169) and (170) are satisfied.

Lemma 25. Given state ρ, bound value c and inverse temperature β, there is a unique partition of the state space that gives the optimal c-bounded work. For $R = \mathrm{rank}(\rho)$ there are at most $R-1$ inequalities that must be checked to determine the partition.

Proof. By 24 we know that $X_\pm < 1/2$, therefore $X_u > 0$. The partition obeys the β-ordering hierarchy (168). Starting from $p(s_1)$ count left to right in $\rho^{\downarrow\beta}$ until you find the furthest $p(s_k)$ s.t. $\sum_{i=1}^{k} p(s) < 1/2$. These will form our trial set for the bounded positive fluctuations $\tilde{\mathcal{X}}_+$. Starting from $p(s_R)$ count from right to left until

you find the furthest $p(s_l)$ s.t. $\sum_{j=l}^{R} p(s_j) < 1/2$. These form our trial set for the bounded negative fluctuations $\tilde{\mathcal{X}}_-$. If this is indeed the correct partition for the state space, the tightest bounds will be (169) on $w(k)$ and (170) on $w(l)$. The inequalities are

$$(1 - \sum_i p(s_i) - \sum_j p(s_j))(w(k) - c) + \sum_i p(s_i)(w(i) - c) + \sum_j p(s_j)(w(j) + c) > W(\rho) \tag{171}$$

$$(1 - \sum_i p(s_i) - \sum_j p(s_j))(w(l) + c) + \sum_i p(s_i)(w(i) - c) + \sum_j p(s_j)(w(j) + c) < W(\rho) \tag{172}$$

Whenever an inequality is satisfied it fixes that fluctuation as being the infinitum of its set. For example, if (171) is satisfied for $w(k)$ then $k \in \tilde{\mathcal{X}}_+$ regardless of $\tilde{\mathcal{X}}_-$. To see this, consider the case that (171) is satisfied for $w(k)$ but (172) isn't for $w(l)$

$$X_u(w(k_-c) + \sum_{i=1}^{k} p(s_i)(w(i) - c) + \sum_{j=l}^{R} p(s_j)(w(j) + c) > W(\rho)$$

$$X_u(w(l) + c) + \sum_{i=1}^{k} p(s_i)(w(i) - c) + \sum_{j=l}^{R} p(s_j)(w(j) + c) > W(\rho)$$

as $w(l)$ doesn't satisfy its bound we have to move further down to $w(l' > l)$. This makes $X_u \to X_u + \sum_{j=l}^{l'} p(s_j)$ and $\sum_{j=l}^{R} p(s_j)(w(j) + c) \to \sum_{j=l}^{R} p(s_j)(w(j) + c) - \sum_{j=l}^{l'} p(s_j)(w(j) + c)$. The left hand side of (171) gains the term

$$\begin{aligned} &+ \sum_{j=l}^{l'} p(s_j)(w(k) - c) - \sum_{j=l}^{l'} p(s_j)(w(j) + c) \\ &= \sum_{j=l}^{l'} p(s_j)(w(k) - w(j)) \end{aligned} \tag{173}$$

As $w(k) \geq w(j)\ \forall\ j = l, \ldots, l'$ this term is positive and the new (171) is guaranteed to be satisfied. Therefore we see that there is a unique partition as satisfying an

inequalities fixes the corresponding (positive or negative) subspace. There are at most $R-1$ inequalities that we need to check, i.e. the "worst case" being when the state space is unbounded and we check all $R-1$ inequalities. □

In summary, the algorithm for determining the partition can be summarised as follows

1. β-order the state $\rho^{\downarrow\beta} = (p(s_1), \ldots, p(s_R))$ where $p(s)^{\beta E_s} \geq p(s_{s+1})^{\beta E_{s+1}}$ and $R = \text{rank}(\rho)$. This ordering gives the unbounded work distribution in descending order $(w(1), \ldots, w(R))$ where $w(s) = \beta^{-1}\log(p(s)e^{\beta E_s}Z')$ and $w(s) \geq w(s+1)$

2. Take the trial partition where you maximise $X_+ = \sum_{i=1}^{k} p(s_i)$ under the condition $X_+ < 1/2$ and $X_- = \sum_{j=l}^{R} p(s_j)$, $X_- < 1/2$. Check inequalities

$$X_u(w(k)-c) + \sum_{i=1}^{k} p(s_i)(w(i)-c) + \sum_{j=l}^{R} p(s_j)(w(j)+c) > W(\rho) \qquad (174)$$

$$X_u(w(l)+c) + \sum_{i=1}^{k} p(s_i)(w(i)-c) + \sum_{j=l}^{R} p(s_j)(w(j)+c) < W(\rho) \qquad (175)$$

3. If (174) is satisfied $p(s_k)$ fixes $X_+ = \sum_{i=1}^{k} p(s_i)$ and similar for (175). Otherwise we perform the next set of inequalities lower in the hierarchy, with $X_+ = \sum_{i=1}^{k-1} p(s_i)$ if (174) is not satisfied and/or $\sum_{j=l+1}^{R} p(s_j)$ if (175) is not satisfied. We repeat this process until we find a pair of inequalities that are simultaneously satisfied, fixing $\tilde{\mathcal{X}}_\pm = \mathcal{X}_\pm$, requiring at most $R-1$ inequalities to be checked.

Case: d=2. This is the simplest case, as either $p(s_1) > 1/2$ or $p(s_2) > 1/2$ or they both $= 1/2$. In the first case we bound the negative fluctuation, and there is a single bound to check

$$w(2) < W^{(\infty)}(\rho) - c \qquad (176)$$

and in the second case we bound the positive fluctuation if

$$w(1) > W^{(\infty)}(\rho) + c \tag{177}$$

and if $p(s_1) = p(s_2) = 1/2$ when the two fluctuations must be equal (as the average of the positive fluctuations = the average of the negative fluctuations) and we can choose to bound one or the other, giving the same free energy.

In the following section we show that the same algorithm is used for determining the partition for the c-bounded work of formation. Note that, unless all fluctuations are unbounded $W^{(c)}(\rho) < W^{(\infty)}(\rho)$ and, as shown in the next section, $W_{\mathrm{F}}^{(c)}(\rho) > W_{\mathrm{F}}^{(\infty)}(\rho)$.

.1.3 Optimal work of formation

Assume there exists some optimal choice of $\{t(s'|s), w(s', s)\}$ that minimize the work cost whilst obeying the c-bounds and microscopic reversibility (5.73). We start in the thermal state with $p(s) = 1/Ze^{-\beta E_s}$ and end the protocol in state ρ with probabilities $q(s')$. The average work is

$$W = \sum_{ss'} \frac{1}{Z} e^{-\beta E_s} t(s'|s) w(s', s) \tag{178}$$

where the $t(s'|s)$ must obey (5.70), (5.71) and (5.72).

Lemma 26. We can always choose new protocol parameters $\tilde{t}(s') = q(s')$ and $\tilde{w}_{s'} = \sum_s \frac{t(s'|s)e^{-\beta E_s}}{p(s')Z} w(s', s)$ that give the same average work as the optimal protocol and obey all the necessary constraints

Proof. Clearly these choices of $\tilde{t}(s')$ satisfy $\sum_{s'} \tilde{t}(s') = 1$ and $\sum_s 1/Ze^{-\beta E_s}\tilde{t}(s') =$

$q(s')$, and the average work is given by

$$\begin{aligned}\tilde{W} &= \sum_{ss} 1/Z e^{-\beta E_s} \tilde{t}(s') \tilde{w}_{s'} = \sum_{s'} \tilde{t}(s') \tilde{w}_{s'} \\ &= \sum_{ss'} p(s') \frac{t(s'|s) e^{-\beta E_s}}{Z p(s')} w(s', s) = W\end{aligned} \tag{179}$$

Note that $\sum_s \frac{t(s'|s)e^{-\beta E_s}}{q(s')Z} = 1$ so $\tilde{w}_{s'}$ is a convex sum of the $w(s', s)$. The reversibility inequality (**??**) demands that

$$\sum_s \tilde{t}(s') e^{\beta(\tilde{w}_{s'} - \Delta E_{ss'})} \leq 1 \tag{180}$$

which we can see to be true as the LHS is

$$\begin{aligned}\text{LHS} &= q(s') e^{\beta(\tilde{w}_{s'} + E_{s'})} \sum_s e^{-\beta E_s} \\ &= q(s') Z e^{\beta(\tilde{w}_{s'} + E_{s'})} \\ &= q(s') Z e^{\beta E_{s'}} e^{\beta \sum_i \frac{t(s'|s)e^{-\beta E_s}}{q(s')Z} w(s',s)} \\ &\leq q(s') Z e^{\beta E_{s'}} \sum_s \frac{t(s'|s) e^{-\beta E_s}}{q(s') Z} e^{\beta w(s',s)} \\ &= \sum_s t(s'|s) e^{\beta(w(s',s) - \Delta E_{ss'})} \leq 1\end{aligned}$$

where in the fourth step we have used the convexity of the exponential function and in the last step we have used the fact that the optimal protocol defined by $\{t(s'|s), w(s', s)\}$ is reversible.

Clearly the new protocol obeys the c-bounds as $\tilde{w}_{s'}$ is a convex combination of $w(s', s)$ and the average of the work distribution is the same for both maps, therefore the spread of $\tilde{w}_{s'}$ about the average is less than or equal to that of $w(s', s)$. □

Substituting $t(s'|s) \to \tilde{t}(s') = q(s')$ and $w(s', s) \to \tilde{w}_{s'}$ into the Lagrangian and

simplifying gives

$$\mathcal{L} = \sum_{s'} f_{s'} w(s') + \sum_{s'} \lambda_{s'} \left(1 - q(s') e^{\beta E_{s'}} Z e^{\beta w(s')}\right) + c \sum_{s'} (\mu_{s'} + \bar{\mu}_{s'})$$

where we have dropped the tilde from $\tilde{w}_{s'}$ and $f_{s'} = q(s')(1+f) - (\bar{\mu}_{s'} - \mu_{s'})$ and $f = \sum_{s'} (\bar{\mu}_{s'} - \mu_{s'})$. $\bar{\mu}$ give the bounds on positive fluctuations and μ give the bounds for negative fluctuations. Maximizing w.r.t $w(s')$ and $\lambda_{s'}$ gives

$$\lambda_{s'} = \max\{ \frac{f_{s'}}{\beta}, 0 \} \tag{181}$$

$$w(s') = -\frac{1}{\beta} \log \left(\frac{\beta \lambda_{s'} q(s') e^{\beta E_{s'}} Z}{f_{s'}} \right) \tag{182}$$

$$= -\frac{1}{\beta} \log \left(q(s') e^{\beta E_{s'}} Z \right) \quad , \quad f_{s'} \neq 0 \tag{183}$$

which is the work value given in the unbounded cases (it contains no $\bar{\mu}_j, \mu_j$ terms). Therefore all fluctuations with $f_{s'} > 0$ are unbounded, with $\mu_{s'}(\bar{\mu}_{s'}) = 0$, giving $f_{s'} = q(s')$ if unbounded or 0 if bounded. The Lagrangian reduces to

$$\mathcal{L} = -\frac{1}{\beta} \sum_{s'} f_{s'} \log \left(q(s') e^{\beta E_{s'}} Z \right) + c \sum_{s'} (\bar{\mu}_{s'} + \mu_{s'}) \tag{184}$$

Which we can now re-write in terms of bounded and unbounded fluctuations. Index all fluctuations with $f_{s'} > 0$ with $s' \in \mathcal{X}_u$, and all those with $f_{s'} = 0$ with $s' \in \mathcal{X}_+$. The Lagrangian becomes

$$\mathcal{L} = -\frac{1}{\beta} \sum_{s' \in \mathcal{X}_u} q(s')(1+f) \log \left(q(s') e^{\beta E_{s'}} Z \right) + c \sum_{s' \in \mathcal{X}_+} (\bar{\mu}_s + \mu_s) \tag{185}$$

$f_s = 0 \; \forall s' \in \mathcal{X}_+$ gives

$$\bar{\mu}_{s'} - \mu_{s'} = q(s')(1+f) \tag{186}$$

Summing over $s \in \mathcal{X}_+$ and solving for $f = \sum_{k'}(\bar{\mu}_{k'} - \mu_{k'})$ gives

$$f = \frac{X_c}{1 - X_c} \tag{187}$$

where $X_c = \sum_{s' \in \mathcal{X}_+} q(s')$. For a given bounded fluctuation only one of the μ Lagrange parameters is non-zero, corresponding to if the fluctuation is positive or negative. Using the above result gives

$$\bar{\mu}_k = p(s_k)(1 + \frac{X_c}{1 - X_c}) = \frac{p(s_k)}{1 - X_c} \tag{188}$$

$$\mu_l = -\frac{p(s_l)}{1 - X_c} \tag{189}$$

The Lagrange parameters are restricted to being positive, $\mu_s \geq 0$ and $\bar{\mu}_s \geq 0 \ \forall \ s$. Therefore as $\mu_s = -p(s)/(1 - X_c)$ which is ≤ 0, therefore all $\mu_s = 0$ and we never bound negative fluctuations.

Using $X_u = 1 - X_c$ we arrive at the c-bounded optimal work of formation

$$W_F^{(c)} = \frac{1}{X_u}\left(-\frac{1}{\beta}\sum_{s' \in \mathcal{X}_u} q(s') \log\left(q(s') e^{\beta E_{s'}} Z\right) + c\, X_c\right) \tag{190}$$

This is exactly the work cost when we calculate the c-bounded work of formation in the following way. Take the optimal unbounded work distribution for forming a state $(w(1), \ldots, w(d))$ with average $W^{(\infty)}(\rho)$. Equation (190) is given by the solution to the equation

$$W = \sum_{s \in \mathcal{X}_u} p(s) w(s) + \sum_{s \in \mathcal{X}_+} p(s)(W + c)$$

$$\therefore W \underbrace{\left(1 - \sum_{s \in \mathcal{X}_+} p(s)\right)}_{X_u} = \sum_{s \in \mathcal{X}_u} p(s) w(s) + c X_+$$

$$\therefore W = \frac{1}{X_u}\left(\sum_{s \in \mathcal{X}_u} p(s) w(s) + c X_+\right)$$

I.e we simply take the optimal unbounded work distribution and if the largest positive fluctuation breaks $w(i) - W \leq c$ we replace it with $W' + c$, recalculating the average each time.

.1.4 Optimal partitions

We now show this algorithm for finding the partition is identical to the algorithm derived in Appendix 1.2 for work extraction, in the case of positive fluctuation bounds only.

Theorem 9. The partition of the state space that gives the c-bounded work of formation is found using the algorithm described in Lemma 25 but bounding only positive fluctuations.

Proof. Take the final state ρ' and β-order it

$$\rho'^{\downarrow\beta} = (p(s_1), \dots p(s_R)) \quad , \quad q(s')e^{\beta E_{s'}} \geq q(s'_{+1})e^{\beta E_{s'+1}} \tag{191}$$

where $R = \text{rank}(\rho)$ (we can discount work values in the unbounded work distribution associated with $p(s) = 0$ as they have no probability of being observed). The unbounded work distribution obeys the inverse β-ordering as $w(s') = -\beta^{-1}\log(q(s')e^{\beta E_{s'}}Z)$, therefore $w(1) \leq w(2) \leq \cdots \leq w(R)$. For a correct partition where $X_+ = \sum_{s'=k}^{R} p(s_i)$ we require that $w(s') \leq W_F(\rho')^{(c)} + c \;\forall\; s < k$. As with the c-bounded work extraction protocol the β-ordering puts these inequalities into a hierarchy, with the bound for $w(k-1)$ being the tightest. As all bounds must be satisfied we need only check the tightest one. For the aforementioned partition the tightest bound is

$$w(k-1) \leq \frac{1}{\sum_{i=k}^{R} p(s_i)} \left(W^{(\infty)} - \sum_{s'=k}^{R} q(s')w(s') + c\sum_{s'=k}^{R} q(s') \right) \tag{192}$$

which can be re-arranged to give

$$(1 - \sum_{s'=k}^{R} q(s'))w(k-1) + \sum_{s'=k}^{R} q(s')w(s') \leq W^{(\infty)} + c \tag{193}$$

which is simply the inequality used for checking a partition of the state space in the c-bounded work extraction protocol, in the case that there are no negative fluctuations that saturate their bounds (174). It can be simplified further to

$$\sum_{s'=1}^{k-1} q(s')(w(k-1) - w(s')) \leq c \tag{194}$$

So the partition is defined by the largest k s.t.

$$\sum_{s'=1}^{k} q(s')(w(k) - w(s')) > c \tag{195}$$

$$\sum_{s'=1}^{k-1} q(s')(w(k-1) - w(s')) \leq c \tag{196}$$

All $w(s)$ with $s \geq k$ are bounded and all with $s < k$ are unbounded. Once again we have at most $R - 1$ inequalities to check. I.e. starting from $k = 2$ we check (195) for $k = 2, 3, \ldots$. □

.1.5 Asymptotics

Theorem 10. $\lim_{c \to 0} W^{(c)}(\rho) = W_{\min}(\rho)$

Proof. As $c \to 0$ the inequalities that determine if a fluctuation saturates its bound (161) and (162) become

$$\frac{1}{\beta} \log \left(p(s) e^{\beta E_s} \right) \geq -\nu/\beta \to s \in \mathcal{X}_+$$

$$\frac{1}{\beta} \log \left(p(s) e^{\beta E_s} \right) \leq -\nu/\beta \to s \in \mathcal{X}_-$$

All fluctuations satisfy one of these bounds, except in the case that $p(s) = 0$ (i.e. there is no fluctuation associated with state $|s\rangle$), therefore $X_u \to 0$. $\gamma = X_u e^{\nu} + Z_+ e^{\beta c} + Z_- e^{-\beta c} \to Z_+ + Z_-$ which can be formulated as

$$\gamma = \sum_s p(s)^0 e^{-\beta E_s} \tag{197}$$

therefore $W^{(c)}(\rho)$ becomes

$$\begin{aligned} \lim_{c\to 0} W^{(c)}(\rho) &= \frac{1}{\beta}\left(\log Z' - \log \sum_s p(s)^0 e^{-\beta E_s}\right) \\ &= W_{\min}(\rho) \end{aligned} \tag{198}$$

□

Theorem 11. $\lim_{c\to 0} W_F^{(c)}(\rho) = W_F^{\min}(\rho)$

Proof. Following the partition algorithm derived in Lemma 9, clearly when $c \to 0$ we bound all but the most negative fluctuation(s), those with value $w(1)$, as this gives the only positively bounded work distribution for which all work values are $\leq W_F^{(0)}$. Any other choice of partition would require both negative and positive bound saturation, which contradicts (188) being positive. Therefore $W_F^{(0)}$ is given by

$$W_F^{(0)} = \min_s \left\{ -\frac{1}{\beta} \log\left(p(s) e^{\beta E'_s} Z\right) \right\} = -\frac{1}{\beta} \log\left(p(s_1) e^{\beta E'_1} Z\right) \tag{199}$$

where $\rho'^{\downarrow\beta} = (p(s_1), \ldots, p(s_R))$. We can interpret

$$w(1) = -\frac{1}{\beta} \log\left(p(s_1) e^{\beta E'_1} Z\right) \tag{200}$$

$-w(1)$ is the "on ramp" (the first segment) of the Lorenz curve of state ρ' (see Figure 5 and [71]). It therefore gives the work that can be extracted from the "thermally sharp state", where all segments either have the same gradient or zero gradient, that just thermomajorizes ρ', see figure 5 below and also [71]. This can

also be interpreted as the upper bound to the work that can be in-deterministically extracted from ρ', and this is precisely the single-shot work of formation. □

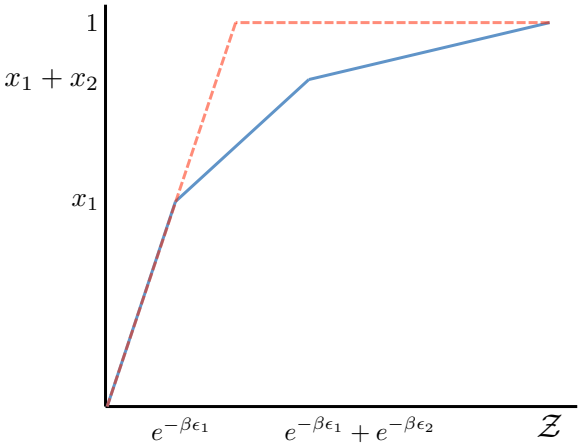


Figure 5: Lorenz curve depicting work of formation of a state. Figure shows the thermally sharp state whose single-shot work content gives the work of formation of state ρ'. This also represents the maximal average work that can be extracted from ρ'

.1.6 Fluctuation bounds

Lemma 27. $W^{(c)}(\rho) \leq W_{\min}(\rho) + c$. $c \in [0, c_{\text{crit}}]$, therefore any gain in work extracted over the single-shot work context of ρ requires fluctuations that are at least as large as the increase in extracted work

Proof. For $c = 0$ we get that $W_{c=0}(\rho) = W_{\min}(\rho)$. $W^{(c)}$ increases as we increase c from 0. We show $W^{(c)}(\rho)$ grows sub-linearly with c, i.e. that

$$\frac{\partial W^{(c)}(\rho)}{\partial c} \leq 1 \quad \forall\, c \tag{201}$$

which implies that $|W^{(c)}(\rho) - W_{\min}(\rho)| \leq c$. Using (158) the derivative of $W^{(c)}(\rho)$ with respect to c is

$$\frac{\partial W^{(c)}(\rho)}{\partial c} = -\frac{1}{\beta\gamma}\frac{\partial \gamma}{\partial c} = \frac{1}{\gamma}\left((X_+ - X_-)e^{\nu} - Z_+ e^{\beta c} + Z_- e^{-\beta c}\right) \tag{202}$$

where $\gamma = X_u e^{\nu} + Z_+ e^{\beta c} + Z_- e^{-\beta c}$. For this to be greater than 1 we require that

$$(X_+ - X_-)e^{\nu} - Z_+ e^{\beta c} + Z_- e^{-\beta c} > X_u e^{\nu} + Z_+ e^{\beta c} + Z_- e^{-\beta c} \tag{203}$$

As $Z_\pm e^{\pm\beta c} \geq 0$ and $e^{\nu} \geq 0$ this cannot be satisfied unless

$$X_+ - X_- > X_u \therefore X_+ > \frac{1}{2} \tag{204}$$

where we have used $X_u = 1 - X_+ - X_-$. We now show that this is never true for c-bounded work distributions. Using $W^{(c)}(\rho) = \sum_s p(s)\theta_s$, where $\theta_s = w(s) - W^{(c)}(\rho)$ is the fluctuation associated with work value $w(s)$, gives $\sum_s p(s)\theta_s = 0$. Divide the fluctuations for the c-bounded protocol θ_s into positive $\theta_\alpha \geq 0$ where $\alpha \in \mathcal{X}_+$ and negative $\theta_\beta < 0$ where $\beta \in \mathcal{X}_-$, giving

$$\sum_\alpha p(s_\alpha)|\theta_\alpha| = \sum_\beta p(s_\beta)|\theta_\beta| \tag{205}$$

according to (135) and (136), bounded fluctuations saturate their bounds. Take the subset of the positive fluctuations that saturate their bounds as $\theta_{\alpha'}$. (205) becomes

$$cX_+ + \sum_{\alpha \neq \alpha'} p(s_\alpha)|\theta_\alpha| = \sum_\beta p(s_\beta)|\theta_\beta| \tag{206}$$

Dividing both sides by $\sum_\beta p(s_\beta)$ gives

$$\frac{\sum_\beta p(s_\beta)|\theta_\beta|}{\sum_\beta p(s_\beta)} = \frac{cX_+ + \sum_{\alpha \neq \alpha'} p(s_\alpha)|\theta_\alpha|}{\sum_\beta p(s_\beta)} \tag{207}$$

the right hand side is strictly larger than c as $X_+ > 1/2$ and therefore $\sum_\beta p(s_\beta) < 1/2$, and $\sum_{\alpha \neq \alpha'} p(s_\alpha)|\theta_\alpha| \geq 0$. The left hand side is a convex sum with all $|\theta_\beta| \geq 0$, therefore at least one $|\theta_\beta|$ must be larger than c. Therefore in a c-bounded protocol $X_+ \leq 1/2$. □

Lemma 28. $\left|W_F^{(c)}(\rho')\right| \geq |W_{\max}(\rho')| - c$. $c \in [0, c_{\text{crit}}]$, therefore any gain in work

extracted over the single-shot work of formation of ρ' requires fluctuations that are at least as large as the increase in extracted work

Proof. if $\left|W_F^{(c)}(\rho')\right| < |W_{\max}(\rho')| - c$ then, taking $w(1) = \max\{w(s')\} = \max\{\beta^{-1}\log q(s'))e^{\beta\epsilon_{s'}}Z\} = |W_{\max}(\rho')|$ when we require that

$$\frac{1}{X_u}\left(\sum_{s'\in\mathcal{X}_u} w(s') + c(1-X_u)\right) + c < w(1) \tag{208}$$

which requires that

$$c < \sum_{s'\in\mathcal{X}_u} (w(1) - w(s')) \tag{209}$$

which contradicts (194) for any partition □

.1.7 c-bounded work and the Jarzinski equality

In this Supplementary Note we show that all c-bounded work distributions obey the Jarzinski equality. The equality is given in the case that a system begins in the thermal state of an initial Hamiltonian H_i and the Hamiltonian is transformed to H_f, causing the state to evolve to a state that is, potentially, out of equilibrium with the final Hamiltonian. The equality is stated as

$$\langle e^{-\beta w}\rangle = e^{-\beta\Delta F} \tag{210}$$

where w is the work random variable (i.e. the work values associated with the protocol), with the convention that they are positive if the work is done on the system, and ΔF is the free energy difference between the equilibrium states of $H_{\mathcal{S}}$ and $H'_{\mathcal{S}}$, given by

$$\Delta F = \frac{1}{\beta}\log\left(\frac{Z}{Z'}\right) \tag{211}$$

Theorem 12. Any protocol that saturates all reversibility constraints immediately

satisfies the Jarzinski equality

Proof. If a protocol, characterised by $\{t(s'|s), w(s', s)\}$, saturates all reversibility constraints then

$$\sum_s t(s'|s)e^{\beta(w(s',s)-E_s+E_{s'})} = 1 \tag{212}$$

Dividing by Z and re-arranging gives

$$\sum_s \frac{e^{-\beta E_s}}{Z} t(s'|s)e^{\beta w(s',s)} = \frac{e^{-\beta E_{s'}}}{Z} \tag{213}$$

if we sum over s', the LHS of this is equivalent to $\langle e^{\beta w(s'|s)} \rangle$ with the initial state being thermal w.r.t $H_{\mathcal{S}}$, $p(s) = Z^{-1}e^{-\beta E_s}$, and the RHS becomes Z'/Z, which is equal to $e^{\beta \Delta F}$. Note that we use the convention of positive work for work extracted from the system, hense our $w(s'|s) = -w(s)$ are the negatives of the work values given in the Jarzinski equality.

□

Theorem 13. The work distributions associated with $W^{(c)}(\rho)$ and $W_F^{(c)}(\rho)$ obey the Jarzinski equality.

First we cover the case of work extraction, $W^{(c)}(\rho)$. The work values are give by

$$w(s) = \frac{1}{\beta} \log(f_s e^{\beta E_s} Z') \tag{214}$$

and the $t(s'|s)$ are given by

$$t(s') = \frac{e^{-\beta E'_{s'}}}{Z'} \tag{215}$$

as derived in Supplementary Note 2. Note that

$$\sum_s f_s = 1 \tag{216}$$

The LHS of the reversibility constraints are

$$\begin{aligned}
&= \sum_s \frac{e^{-\beta E_{s'}}}{Z'} e^{\beta(w(s)-E_s+E_{s'})} \\
&= \sum_s \frac{1}{Z'} e^{\log(f_s Z'))} \\
&= \sum_s f_s = 1
\end{aligned}$$

and can similarly be shown for $W^{(c)}(\rho)$.

.2 c-bounded Carnot efficiency

In this Appendix we calculate the c-bounded Carnot efficiency for a single qubit quantum engine. The engine operates by moving a qubit $\rho = x\,|0\rangle\langle 0| + (1-x)\,|1\rangle\langle 1|$ with Hamiltonian $H_s = E\,|0\rangle\langle 0|$ between two baths with inverse temperature β_H and β_C, with $\beta_H < \beta_C$. The engine cycle begins with the qubit in thermal equilibrium with the cold bath. It is then placed in contact with the hot bath, and extracting work by allowing the state to equilibrate. The final step in the cycle is to return the qubit to the cold bath, allowing it to equilibrate and extracting work.

Case: $\mathbf{c \to \infty}$. Following the proof given in [124], we show that in the case that fluctuations are unbounded it is possible to reach Carnot efficiency. The equilibrium state of the qubit if given by

$$\rho_{H,C} = \frac{1}{Z_{H,C}} e^{-\beta_{H,C} E}\,|0\rangle\langle 0| + \frac{1}{Z_{H,C}}\,|1\rangle\langle 1| \tag{217}$$

where $Z_{H,C} = e^{-\beta_{H,C} E} + 1$. When ρ_C is placed in thermal contact with the hot bath we can extract a maximal average work given by the free energy difference

$$\begin{aligned}
F(\rho_C, \beta_H) - F(\rho_H, \beta_H) = \frac{1}{\beta_H} \log\left(\frac{Z_H}{Z_C}\right) \\
+ \operatorname{tr}(H_s \rho_c) \left(\frac{\beta_H - \beta_C}{\beta_H}\right)
\end{aligned} \tag{218}$$

where $\mathrm{tr}\left(H_s\rho_C\right) = Z_C^{-1}e^{-\beta_C E}E$ and work values

$$\begin{aligned} w_{0,H} &= \frac{1}{\beta_H}\log\left(\frac{Z_H}{Z_C}e^{E(\beta_H-\beta_C)}\right) \\ w_{1,H} &= \frac{1}{\beta_H}\log\left(\frac{Z_H}{Z_C}\right) \end{aligned} \tag{219}$$

where we have used the result for the optimal unbounded work extraction protocol $w(s) = \beta^{-1}\log p(s)e^{\beta E_s}Z$. Returning the qubit to the cold bath we extract work

$$\begin{aligned} F(\rho_H,\beta_C) - F(\rho_C,\beta_C) = \frac{1}{\beta_C}\log\left(\frac{Z_C}{Z_H}\right) \\ + \mathrm{tr}\left(H_s\rho_H\right)\left(\frac{\beta_C-\beta_H}{\beta_C}\right) \end{aligned} \tag{220}$$

with work values

$$\begin{aligned} w_{0,C} &= \frac{1}{\beta_C}\log\left(\frac{Z_C}{Z_H}e^{E(\beta_C-\beta_H)}\right) \\ w_{1,C} &= \frac{1}{\beta_C}\log\left(\frac{Z_C}{Z_H}\right) \end{aligned} \tag{221}$$

where $\mathrm{tr}\left(H_s\rho_H\right) = Z_H^{-1}e^{-\beta_H E}E$. The total work extracted in one cycle is

$$W_{\mathrm{tot}} = \left(\frac{1}{\beta_H} - \frac{1}{\beta_C}\right)(S_H - S_C) \tag{222}$$

where $S_{H,C} = -\log Z_{H,C} - \beta_{H,C}\mathrm{tr}\left(H_s\rho_{H,C}\right)$ is the Von Neumann entropy of the Gibbs state with inverse temperature $\beta_{H,C}$. Applying the first law of thermodynamics $\Delta U = Q - W$ to the first step (extracting work from the hot bath), we get

$$\mathrm{tr}\left(H_s\rho_H\right) - \mathrm{tr}\left(H_s\rho_C\right) = Q_H - \left(F(\rho_C,\beta_H) - F(\rho_H,\beta_H)\right) \tag{223}$$

which simplifies to

$$Q_H = \frac{1}{\beta_H}\left(S_H - S_C\right) \tag{224}$$

where Q_H is the heat flow out of the hot bath. The efficiency is given by the ratio of the total work to Q_H, giving

$$\frac{W_{\text{tot}}}{Q_H} = 1 - \frac{\beta_H}{\beta_C} \tag{225}$$

which is the Carnot efficiency.

Case: c finite.

$E > 0$ implies that $x < 1/2$ for all thermal states of ρ, therefore by the partitioning algorithm derived in Supplementary Note 4, the least likely fluctuation $w_0^{(c)}$ associated with subspace $|0\rangle$ is the only fluctuation that can be bounded. Using the result of Supplementary note 3, the c-bounded work content for a qubit $\rho^{\downarrow\beta} = (x_1, x_2)$ is given by

$$W^{(c)}(\rho) = \frac{1}{\beta}\log Z + \begin{cases} F(\rho)\,, & x_1 e^{\beta E_1} \leq e^{\beta(F(\rho)+c)} \\ & x_2 e^{\beta E_2} \geq e^{\beta(F(\rho)-c)} \\ F_1^{(c)}(\rho)\,, & w(0) < W(\rho) - c \\ F_2^{(c)}(\rho)\,, & w(0) > W(\rho) + c \end{cases} \tag{226}$$

where $F_1^{(c)}(\rho) = -\beta^{-1}\log\left(e^{-c\beta+\frac{c\beta}{1-x}} + e^{-c\beta-E\beta}\right)$ and $F_2^{(c)}(\rho) = -\beta^{-1}\log\left(e^{c\beta-\frac{c\beta}{1-x}} + e^{c\beta-E\beta}\right)$, $W(\rho)$ is the unbounded work given by the free energy, and $w(0) = \beta^{-1}\log(xe^{\beta E}Z)$. In the first step, putting the qubit in contact with the hot bath, the work we extract is bounded negatively if $w_{0,H} < F(\rho_C, \beta_H) - F(\rho_H, \beta_H) - c$, which gives the inequality

$$\frac{E}{Z_C}\left(\frac{\beta_C - \beta_H}{\beta_H}\right) > c \tag{227}$$

note that $\beta_C > \beta_H$ therefore the LHS of (227) is positive. In order to break the positive fluctuation bound we would require -LHS $> c$ which is impossible for positive c, therefore in the first part of the engine cycle the work extracted is either unbounded or negatively bounded. For the second cycle, re-equilibriating the qubit with the cold bath, the inequality that implies a positively bounded work is given

by $w_{0,C} > F(\rho_H, \beta_C) - F(\rho_C, \beta_C) + c$, which simplifies to

$$\frac{E}{Z_H}\left(\frac{\beta_C - \beta_H}{\beta_C}\right) > c \tag{228}$$

Again the LHS is positive, and we would require -LHS $> c$ in the case of a negatively bounded protocol. Therefore on the second part of the cycle the work is either unbounded or positively bounded. Also note that as $\beta_H < \beta_C$ and therefore $Z_C < Z_H$, satisfying inequality (227) implies inequality (228) is also satisfied. There are therefore three cases, 1) the work is unbounded and we can achieve Carnot efficiency, 2) the work extracted from the hot bath is negatively bounded, and 3) 2 and the work extracted from the cold bath is positively bounded.

the c-bounded work extracted from the hot bath, in the case that inequality (227) is satisfied, is given by

$$W_1^{(c)} = \frac{1}{\beta_H}\log\left(\frac{Z_H}{e^{c\beta_H Z_C} + e^{-\beta_H E}}\right) + c \tag{229}$$

where we have used the expressions in (226). If inequality (228) is satisfied for the second part of the cycle, then we get

$$W_2^{(c)} = \frac{1}{\beta_C}\log\left(\frac{Z_C}{e^{-c\beta_C Z_H} + e^{-\beta_C E}}\right) - c \tag{230}$$

Using $Q_H = \Delta U + W_1^{(c)}$, the efficiency is given by

$$\eta_1^{(c)} = \frac{W_1^{(c)} + F(\rho_H, \beta_C) - F(\rho_C, \beta_C)}{\Delta U + W_1^{(c)}} \tag{231}$$

if inequality (227) is satisfied and

$$\eta_2^{(c)} = \frac{W_1^{(c)} + W_2^{(c)}}{\Delta U + W_1^{(c)}} \tag{232}$$

if inequalities (227) and (228) are satisfied. As the temperature difference between

hot and cold baths increases, the efficiency increases. In the unbounded case, the Carnot efficiency is bounded from above by 1. As $\beta_H/\beta_C \to 0$, inequality (228) cannot be satisfied for any finite c. To find the upper bound of $\eta_1^{(c)}$ we make the change of variables $\beta_C/n = \beta_H$, i.e. $T_H = nT_C$ and n is the ratio of the two bath temperatures. We then take $\beta_H \to \infty$ and $n \to \infty$.

$$\lim_{n\to 0}\left(\lim_{\beta_H\to\infty}\left(\eta_1^{(c)}\right)\right) = 1 - \frac{E}{2(2c+E)} \tag{233}$$

Of course as $E \to 0$ the efficiency tends to 1, but this is because we are no longer extracting any work (as the two Gibbs states are identical). As $c \to \infty$ we recover the Carnot efficiency upper bound, although for any finite c we can never reach an efficiency of 1. Therefore all realistic engines of this form have an upper bound to their efficiency defined by their energy gap and fragility c. This represents a general physical upper bound on the efficiency of a the qubit engine, given by its ability to withstand fluctuations. Also notice that as $c \to 0$ the maximal efficiency is bounded from below by 1/2. For $c = 0$ the efficiency is zero, as no work can be extracted. Therefore we observe, in this model at least, a genuine discontinuity between the capabilities of a thermal machine in the single-shot regime. Allowing for arbitrarily small fluctuations in principle will allow the engine to reach a maximum efficiency that is bounded from below by 1/2. But if we demand that the work is deterministic, the efficiency is zero.

.3 Finite bath second law

In this appendix, we derive the finite bath second law, and related results

.3.1 finite bath thermal operations

In this appendix we derive Theorem 8

First we address the case of infinite bath ($\gamma = 0$).

Lemma 29. When $\gamma = 0$ all thermal operations saturating the inequality $\langle W \rangle \geq -\Delta F$ are of the form

$$t(E's'|Es) = f(E's's)\delta(E - E' - f_{s'} + f_s)$$

where $f(E's's)$ obeys

$$\sum_{s'} f(E - f_{s'} + f_s, s', s) = 1 \quad \forall\ s \tag{234}$$

$$\sum_{s} P(s) f(E's's) = P(s') \quad \forall\ s' \tag{235}$$

$$0 \leq f(E's's) \leq 1 \quad \forall\ E's's \tag{236}$$

where $f_s = -\beta^{-1} \log P(s)$ is the fine-grained entropy.

Proof. In order to achieve $W = -\Delta F$, a thermal operation $t(E's'|Es)$ must saturate the upper bound derived in Theorem 1, and the right hand side of the bound must be equal to $-\Delta F$. First note that in Theorem 1 the optimal $Q(E's's)$ are given uniquely by

$$Q(E's's) = \frac{1}{\beta} \log \left(\frac{P(s) p_b(E')}{p(E's')} \right) \ . \tag{237}$$

By Theorem **??**, as the relative entropy is a distance measure for probability distributions the upper bound is $-\Delta F$ if and only if $P(E's') = p_b(E')P(s')$ and therefore, for all optimal thermal operations

$$Q(E's's) = \frac{1}{\beta} \log \left(\frac{P(s)}{P(s')} \right) = f_s - f_{s'} \ . \tag{238}$$

Finally, note that the thermal operation only construct an optimal Q if the re-

versibility constraint (5.40) is saturated. This requires that

$$\sum_{Es} \frac{R(E's'|s)}{p_b(E')} t(E's'|Es) \frac{p_b(E)}{R(E's'|s)} e^{\beta(E-E')} = \sum_s \frac{R(E's'|s)}{p_b(E')} \exp\left[\beta \frac{\sum_E p_b(E) t(E's'|Es)(E-E')}{R(E's'|s)}\right] \tag{239}$$

Therefore the sum over E in the exponent must have only a single term by convexity, and therefore $t(E's'|Es)$ be a delta function killing the sum over E. This, combined with the definition of $Q(E's's)$ in (**??**) implies that $t(E's'|Es)$ is of the form

$$t(E's'|Es) = f(E's's)\delta(E - E' - f_{s'} + f_s) \tag{240}$$

where $f(E's's)$ is a function that, by substituting (29) into (169),(170),(171) obeys

$$\sum_{s'} f(E - f_{s'} + f_s, s', s) = 1 \quad \forall\ s \tag{241}$$

$$\sum_s P(s) f(E's's) p_b(E' + f_{s'} - f_s) = P(s') p_b(E') \quad \forall\ s', E' \tag{242}$$

$$0 \leq f(E's's) \leq 1 \quad \forall\ E's's \tag{243}$$

In the limit $\gamma \to 0$ the distribution $p_b(E)$ tends to a constant. Hence, equality (242) becomes (235). □

As we now know the form of any optimal thermal operation to zero order in γ, we can therefore express any optimal thermal operation to order γ as a perturbation

$$t(E's'|Es) = f(E's's)\delta(E - E' - f_{s'} - f_s) + \gamma t_1(E's'|Es)$$

Substituting this into (169-171) and using (234-236) we get the following constraints for the general correction to the map

$$\sum_{E's'} t_1(E's'|s) = 0 \quad \forall\ s \tag{244}$$

$$\sum_{s} P(s)t_1(E's'|s) = 0 \quad \forall\ s' \tag{245}$$

$$\sum_{E's'} t_1(E's'|s) = 0 \quad \forall\ s \tag{246}$$

$$\sum_{E',s,E} P(s)t_1(E's'|s) = 0 \quad \forall\ s' \tag{247}$$

Using this we now find an expression for the optimal work in terms of $f(E's's)$ only.

Lemma 30. For any given transformation $P(s) \to P(s')$ there is a map $t(E's'|Es)$ which extracts work

$$W = -\Delta F - \frac{\gamma^2}{2\beta} \sum_{E's'} P(s')p_b(E')E'^2 \left(\frac{\sum_s P(s)f(E's's)(f_{s'} - f_s)}{P(s')} \right)^2 + \mathcal{O}(\gamma^2) \ ,$$

up to first order in γ. (Recall that $p_b(E')E'^2$ is of order $1/\gamma$).

Proof. The upper bound to the work in Theorem 1 can be written as

$$W \leq -\Delta F - \beta^{-1} \sum_{E's'} P(E', s') \log \left(\frac{P(E', s')}{p_b(E')P(s')} \right)$$

. To first order in γ, optimal thermal operations are of the form

$$t(E's'|Es) = f(E's's)\delta(E - E' - f_{s'} + f_s) + \gamma t_1(E's'|Es) \ .$$

Let us calculate the joint final probability distribution $P(E's')$ to first order in γ.

First,

$$P(E's') = \sum_{Es} P(s)p_b(E)t(E's'|Es)$$
$$= \sum_{s} P(s)p_b(E' + f_{s'} - f_s)f(E's's) + \gamma \sum_{s} P(s)p_b(E')t_1(E's'|Es) \ .$$

Second, we expand $p_b(E' + f_{s'} - f_s)$ to first order in γ

$$p_b(E' + f_{s'} - f_s) = p_b(E')(1 - \gamma E'(f_{s'} - f_s) + \gamma/2(E'^2\gamma - 1)(f_{s'} - f_s)^2) + \mathcal{O}(\gamma^{3/2}) \quad (248)$$

Note we have taken $E' \sim \gamma^{-1/2}$ as $\sum_{E'} p_b(E')E'^2 = \gamma^{-1}$. Therefore, to first order in γ, we recover the joint final probability distribution

$$\begin{aligned} P(E's') &= \sum_{s} P(s)p_b(E')f(E's's)\left(1 - \gamma E'(f_{s'} - f_s) + \gamma/2(E'^2\gamma - 1)(f_{s'} - f_s)^2\right) \\ &+ \gamma \sum_{s} P(s)p_b(E')t_1(E's'|s) + \mathcal{O}(\gamma^{3/2}) \end{aligned} \quad (249)$$

and therefore

$$\frac{P(E's')}{P(s')p_b(E')} - 1 = \underbrace{-\frac{\gamma E'}{P(s')} \sum_{s} P(s)f(E's's)(f_{s'} - f_s)}_{[A]} \quad (250)$$
$$+ \underbrace{\frac{\gamma}{2P(s')}(E'^2\gamma - 1) \sum_{s} P(s)f(E's's)(f_{s'} - f_s)^2}_{[B]} \quad (251)$$
$$+ \underbrace{\frac{\gamma}{P(s')p_b(E')} \sum_{sE} P(s)p_b(E)t_1(E's'|Es)}_{[C]} \quad (252)$$

where we have used (241).

Define

$$x(E's') = \frac{P(E's')}{p_b(E')P(s')} - 1 \ , \quad (253)$$

and note that $x(E's') \sim \mathcal{O}(\gamma)$. Using $\log(1 + x(E's')) = x - 1/2x^2 + \mathcal{O}(x^3)$ as $x \ll 1$ we can expand equation (5.48) to first order in $x(E's')$, giving a correction to the

free energy of

$$-\frac{1}{\beta}\sum_{E's'}P(E's')\log\left(\frac{P(E's')}{p_b(E')P(s')}\right) = -\frac{1}{\beta}\sum_{E's'}P(E's')\left(\frac{P(E',s')}{p_b(E')P(s')}-1\right) + \frac{1}{\beta}\sum_{E's'}P(E's')\left(\frac{P(E's')}{p_b(E')P(s')}-1\right)^2+\mathcal{O}(\gamma^2) \quad (254)$$

We now substitute here to first order in γ. First, we show that the contribution from term $[C]$ in (252) is zero, and hence we can choose an optimal thermal operation whereby E is specified by E', s', s and the upper bound to work is saturated. As term $[C]$ is $\mathcal{O}(\gamma)$, its contribution to the work appears in the $\mathcal{O}(x)$ term in (254) only

$$-\frac{1}{\beta}\sum_{E's'}P(E's')\left(\frac{P(E',s')}{p_b(E')P(s')}-1\right) \quad (255)$$

To zero'th order in γ, $P(E's') = \sum_s P(s)f(E's's)p_b(E') = P(s')p_b(E')$. Substituting this and $[C]$ into the above expression gives

$$-\frac{1}{\beta}\sum_{E's'}p_b(E')P(s')\frac{\gamma}{P(s')p_b(E')}\sum_{sE}P(s)p_b(E)t_1(E's'|Es) = -\frac{\gamma}{\beta}\sum_{E's'Es}P(s)p_b(E)t_1(E's'|Es) \quad (256)$$

Applying constraints (241) and (242) sets this term to zero. Substituting $[A]$ and $[B]$ into (254) and working to $\mathcal{O}(\gamma)$ gives

$$\begin{aligned} W &= -\Delta F - \frac{1}{\beta}\sum_{E's'}P(E's')\left(\frac{P(E',s')}{p_b(E')P(s')}-1\right) + \frac{1}{\beta}\sum_{E's'}P(E's')\left(\frac{P(E's')}{p_b(E')P(s')}-1\right)^2+\mathcal{O}(\gamma^2) \\ &= -\Delta F - \frac{\gamma^2}{\beta}\sum_{E's'}P(s')p_b(E')E'^2\left(\frac{\sum_s P(s)f(E's's)(f_{s'}-f_s)}{P(s')}\right)^2 \\ &\quad - \frac{\gamma}{2\beta}\sum_{E's'}p_b(E')P(s')(E'^2\gamma-1)\sum_s\frac{P(s)f(E's's)(f_{s'}-f_s)^2}{P(s')} \end{aligned}$$

where we have used (241), (242), (248) and $E' \sim \mathcal{O}(\gamma^{-1/2})$ as $\langle E'^2\rangle = \gamma^{-1}$.

We now show that the second term is equal to zero to first order in γ

$$\frac{\gamma}{2\beta}\sum_{E's'}p_b(E')P(s')(E'^2\gamma-1)\sum_s\frac{P(s)f(E's's)(f_{s'}-f_s)^2}{P(s')}=0 \tag{257}$$

First note that $\sum_{EE'ss'}P(s)p_b(E')t(E's'|Es)=1$ and therefore

$$1=\sum_{E'ss'}P(s)p_b(E'-f_{s'}+f_s)f(E's's)+\gamma\underbrace{\sum_{EE'ss'}P(s)p_b(E)t_1(E's'|Es)}_{=0\text{ by (246) and (247)}} \tag{258}$$

Expanding $p_b(E'-f_{s'}+f_s)$ gives

$$\begin{aligned}1&=\sum_{E's's}P(s)p_b(E')(1-E'\gamma(f_{s'}-f_s)+\frac{\gamma}{2}(E'^2\gamma-1)(f_{s'}-f_s)^2)f(E's's)+\mathcal{O}(\gamma^2)\\&=1-\gamma\sum_{E's's}P(s)p_b(E')E'f(E's's)(f_{s'}-f_s)\\&+\frac{\gamma}{2}\sum_{E's's}P(s)p_b(E')(E'^2\gamma-1)(f_{s'}-f_s)^2)f(E's's)+\mathcal{O}(\gamma^2)\end{aligned}$$

To see that the second term on the right hand side is zero, note that for $\sum_{E's's}P(s)p_b(E')(E'^2\gamma-1)f_{s'}$ (241) implies $\sum_s P(s)f(E's's)=P(s')\ \forall\ E'$, removing the E' dependence from the $f(E's's)$. Then using $\langle E'^2\rangle=\gamma^{-1}$ sets this to zero. Similarly, with (242) the term $-\sum_{E's's}P(s)p_b(E')(E'^2\gamma-1)f_s$ is identical to zero. Therefore we recover that

$$\frac{\gamma}{2}\sum_{E's's}P(s)p_b(E')(E'^2\gamma-1)(f_{s'}-f_s)^2)f(E's's)=\mathcal{O}(\gamma^2) \tag{259}$$

□

Theorem 8 *In any process, work is bounded form above by*

$$W\leq-\Delta F-\frac{1}{2\beta C}\Delta S^2\ .$$

Proof. Starting with the equation for the optimal work derived in Lemma 30

$$W = -\Delta F - \underbrace{\frac{\gamma^2}{2\beta}\sum_{E's'} P(s')p_b(E')E'^2\left(\frac{\sum_s P(s)f(E's's)(f_{s'}-f_s)}{P(s')}\right)^2}_{[A]}$$
$$\underbrace{\frac{\gamma}{2\beta}\sum_{E's's} p_b(E')(E'^2\gamma-1)P(s)f(E's's)(f_{s'}-f_s)^2}_{[B]} \tag{260}$$

the upper bound can be derived through simple convexity arguments. First consider term $[A]$.

$$\begin{aligned}
[A] &= -\frac{\gamma^2}{2\beta}\sum_{E's'} P(s')p_b(E')E'^2\left(\frac{\sum_s P(s)f(E's's)(f_{s'}-f_s)}{P(s')}\right)^2 \\
&\leq -\frac{\gamma}{2\beta}\sum_{E'} p_b(E')E'^2\left(\frac{\sum_{ss'} P(s)f(E's's)(f_{s'}-f_s)}{P(s')}\right)^2 \\
&= -\frac{\gamma}{2\beta^3}\Delta S^2
\end{aligned}$$

where we have used (241) and (242) and the concavity of the function $g(x) = -x^2$. Finally, consider term $[B]$. By (241) and (242) the sum $\sum_{ss'} P(s)f(E's's)$ is convex positive for all E', and therefore

$$\begin{aligned}
[B] &\leq \frac{\gamma}{2\beta}\sum_{E'} p_b(E')(E'^2\gamma-1)(\sum_{ss'} P(s)f(E's's)(f_{s'}-f_s))^2 \\
&= \frac{\gamma}{2\beta^3}\sum_{E'} p_b(E')(E'^2\gamma-1)\Delta S^2 \\
&= 0
\end{aligned}$$

where we have used (241) and (242), the convexity of the function $g(x) = x^2$ and $\langle E'^2\rangle = \gamma^{-1}$. Using these upper bounds for terms [A] and [B], and $\gamma = \beta^2/C$, we construct the desired upper bound for the work associated with the thermal operation. □

.4 Permissions

The results presented in this thesis have permission to be replublished. Permission for the articles [114] and [4] follow from these publication being open access. Permission for articles [115, 113] are given in the terms of service of the APS Physical Review journals.

Bibliography

[1] Scott Aaronson. Is quantum mechanics an island in theoryspace? *arXiv preprint quant-ph/0401062*, 2004.

[2] Johan Åberg. Truly work-like work extraction via a single-shot analysis. *Nature communications*, 4:1925, 2013.

[3] Johan Åberg. Fully quantum fluctuation theorems. *Physical Review X*, 8(1):011019, 2018.

[4] Sabri W Al-Safi and Jonathan Richens. Reversibility and the structure of the local state space. *New Journal of Physics*, 17(12):123001, 2015.

[5] Sabri W Al-Safi and Anthony J Short. Reversible dynamics in strongly nonlocal boxworld systems. *Journal of Physics A: Mathematical and Theoretical*, 47(32):325303, 2014.

[6] AE Allahverdyan and Th M Nieuwenhuizen. Fluctuations of work from quantum subensembles: The case against quantum work-fluctuation theorems. *Physical Review E*, 71(6):066102, 2005.

[7] Mateus Araújo, Cyril Branciard, Fabio Costa, Adrien Feix, Christina Giarmatzi, and Časlav Brukner. Witnessing causal nonseparability. *New Journal of Physics*, 17(10):102001, 2015.

[8] Mateus Araújo, Adrien Feix, Miguel Navascués, and Časlav Brukner. A purification postulate for quantum mechanics with indefinite causal order. *Quantum*, 1:10, 2017.

[9] Miriam Backens. The zx-calculus is complete for stabilizer quantum mechanics. *New Journal of Physics*, 16(9):093021, 2014.

[10] John C Baez. Division algebras and quantum theory. *Foundations of Physics*, 42(7):819–855, 2012.

[11] George Phillip Barker. Theory of cones. *Linear Algebra and its Applications*, 39:263–291, 1981.

[12] Howard Barnum, Jonathan Barrett, Matthew Leifer, and Alexander Wilce. Generalized no-broadcasting theorem. *Physical review letters*, 99(24):240501, 2007.

[13] Howard Barnum, Ciarán M Lee, Carlo Maria Scandolo, and John H Selby. Ruling out higher-order interference from purity principles. *Entropy*, 19(6):253, 2017.

[14] Howard Barnum, Markus P Müller, and Cozmin Ududec. Higher-order interference and single-system postulates characterizing quantum theory. *New Journal of Physics*, 16(12):123029, 2014.

[15] Jonathan Barrett. Information processing in generalized probabilistic theories. *Physical Review A*, 75(3):032304, 2007.

[16] Alessio Belenchia, Giacomo Mauro D'Ariano, and Paolo Perinotti. Universality of computation in real quantum theory. *EPL (Europhysics Letters)*, 104(2):20006, 2013.

[17] John S Bell. Einstein-podolsky-rosen experiments. In *John S Bell on the Foundations of Quantum Mechanics*, pages 74–83. World Scientific, 2001.

[18] Ingemar Bengtsson, Stephan Weis, and Karol Życzkowski. Geometry of the set of mixed quantum states: An apophatic approach. In *Geometric Methods in Physics*, pages 175–197. Springer, 2013.

[19] Ingemar Bengtsson and Karol Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge university press, 2017.

[20] Charles H Bennett. Notes on landauer's principle, reversible computation, and maxwell's demon. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 34(3):501–510, 2003.

[21] P Bocchieri and A Loinger. Quantum recurrence theorem. *Physical Review*, 107(2):337, 1957.

[22] Max Born, Werner Heisenberg, and Pascual Jordan. Zur quantenmechanik. ii. *Zeitschrift für Physik*, 35(8-9):557–615, 1926.

[23] Cyril Branciard, Mateus Araújo, Adrien Feix, Fabio Costa, and Časlav Brukner. The simplest causal inequalities and their violation. *New Journal of Physics*, 18(1):013008, 2015.

[24] Fernando Brandao, Michał Horodecki, Nelly Ng, Jonathan Oppenheim, and Stephanie Wehner. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences*, 112(11):3275–3279, 2015.

[25] Časlav Brukner. Quantum causality. *Nature Physics*, 10(4):259, 2014.

[26] Časlav Brukner, Marek Żukowski, Jian-Wei Pan, and Anton Zeilinger. Bellâs inequalities and quantum communication complexity. *Physical review letters*, 92(12):127901, 2004.

[27] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86(2):419, 2014.

[28] Francesco Buscemi. All entangled quantum states are nonlocal. *Physical review letters*, 108(20):200401, 2012.

[29] Adán Cabello. Interpretations of quantum theory: a map of madness, 2015.

[30] Michele Campisi, Peter Hänggi, and Peter Talkner. Colloquium: Quantum fluctuation relations: Foundations and applications. *Reviews of Modern Physics*, 83(3):771, 2011.

[31] Michele Campisi, Jukka Pekola, and Rosario Fazio. Nonequilibrium fluctuations in quantum heat engines: theory, example, and possible solid state experiments. *New Journal of Physics*, 17(3):035012, 2015.

[32] Giulio Chiribella, Giacomo Mauro DâAriano, and Paolo Perinotti. Probabilistic theories with purification. *Physical Review A*, 81(6):062348, 2010.

[33] Giulio Chiribella, Giacomo Mauro DâAriano, and Paolo Perinotti. Informational derivation of quantum theory. *Physical Review A*, 84(1):012311, 2011.

[34] Giulio Chiribella, Giacomo Mauro DâAriano, and Paolo Perinotti. Quantum theory, namely the pure and reversible theory of information. *Entropy*, 14(10):1877–1893, 2012.

[35] Giulio Chiribella, Giacomo Mauro DâAriano, Paolo Perinotti, and Benoit Valiron. Quantum computations without definite causal structure. *Physical Review A*, 88(2):022318, 2013.

[36] Giulio Chiribella and Carlo Maria Scandolo. Conservation of information and the foundations of quantum mechanics. In *EPJ Web of Conferences*, volume 95, page 03003. EDP Sciences, 2015.

[37] Giulio Chiribella and Carlo Maria Scandolo. Operational axioms for diagonalizing states. *arXiv preprint arXiv:1506.00380*, 2015.

[38] Boris S Cirel'son. Quantum generalizations of bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980.

[39] John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. Proposed experiment to test local hidden-variable theories. *Physical review letters*, 23(15):880, 1969.

[40] Bob Coecke and Jonathan Barrett. *Bounds on computation from physical principles*. PhD thesis, University of Oxford, 2017.

[41] Piotr Ćwikliński, Michał Studziński, Michał Horodecki, and Jonathan Oppenheim. Limitations on the evolution of quantum coherences: towards fully quantum second laws of thermodynamics. *Physical review letters*, 115(21):210403, 2015.

[42] Oscar CO Dahlsten, Renato Renner, Elisabeth Rieper, and Vlatko Vedral. Inadequacy of von neumann entropy for characterizing extractable work. *New Journal of Physics*, 13(5):053015, 2011.

[43] B Dakić, Tomasz Paterek, and Č Brukner. Density cubes and higher-order interference theories. *New Journal of Physics*, 16(2):023028, 2014.

[44] Sorin Dascalescu, Constantin Nastasescu, and Serban Raianu. *Hopf algebra: An introduction*. CRC Press, 2000.

[45] Gonzalo de la Torre, Lluís Masanes, Anthony J Short, and Markus P Müller. Deriving quantum theory from its local structure and reversibility. *Physical review letters*, 109(9):090403, 2012.

[46] GM DÊ$\frac{1}{4}$Ariano, Franco Manessi, and Paolo Perinotti. Determinism without causality. *Physica Scripta*, 2014(T163):014013, 2014.

[47] Giacomo Mauro DâAriano. Causality re-established. *Phil. Trans. R. Soc. A*, 376(2123):20170313, 2018.

[48] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935.

[49] Massimiliano Esposito, Katja Lindenberg, and Christian Van den Broeck. Thermoelectric efficiency at maximum power in a quantum dot. *EPL (Europhysics Letters)*, 85(6):60010, 2009.

[50] Massimiliano Esposito, Katja Lindenberg, and Christian Van den Broeck. Entropy production as correlation between system and reservoir. *New Journal of Physics*, 12(1):013013, 2010.

[51] Tobias Fritz. Beyond bellâs theorem ii: Scenarios with arbitrary causal structure. *Communications in Mathematical Physics*, 341(2):391–434, 2016.

[52] Ken Funo, Yûto Murashita, and Masahito Ueda. Quantum nonequilibrium equalities with absolute irreversibility. *New Journal of Physics*, 17(7):075005, 2015.

[53] Thomas D Galley and Lluis Masanes. Impossibility of mixed-state purification in any alternative to the born rule. *arXiv preprint arXiv:1801.06414*, 2018.

[54] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Physical review letters*, 96(1):010401, 2006.

[55] Alison Gopnik, David M Sobel, Laura E Schulz, and Clark Glymour. Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental psychology*, 37(5):620, 2001.

[56] Gilad Gour, Markus P Müller, Varun Narasimhachar, Robert W Spekkens, and Nicole Yunger Halpern. The resource theory of informational nonequilibrium in thermodynamics. *Physics Reports*, 583:1–58, 2015.

[57] David Gross, Markus Müller, Roger Colbeck, and Oscar CO Dahlsten. All reversible dynamics in maximally nonlocal theories are trivial. *Physical review letters*, 104(8):080402, 2010.

[58] Branko Grünbaum, Victor Klee, Micha A Perles, and Geoffrey Colin Shephard. Convex polytopes. 1967.

[59] Yelena Guryanova, Ralph Silva, Anthony J Short, Paul Skrzypczyk, Nicolas Brunner, and Sandu Popescu. Exploring the limits of no backward in time signalling. *arXiv preprint arXiv:1708.00669*, 2017.

[60] Paul R Halmos. *Finite-dimensional vector spaces*. Courier Dover Publications, 2017.

[61] Nicole Yunger Halpern. Beyond heat baths ii: Framework for generalized thermodynamic resource theories. *Journal of Physics A: Mathematical and Theoretical*, 2018.

[62] Nicole Yunger Halpern, Philippe Faist, Jonathan Oppenheim, and Andreas Winter. Microcanonical and resource-theoretic derivations of the thermal state of a quantum system with noncommuting charges. *Nature communications*, 7:12051, 2016.

[63] Nicole Yunger Halpern, Andrew JP Garner, Oscar CO Dahlsten, and Vlatko Vedral. Introducing one-shot work into fluctuation relations. *New Journal of Physics*, 17(9):095003, 2015.

[64] Nicole Yunger Halpern and Joseph M Renes. Beyond heat baths: Generalized resource theories for small-scale thermodynamics. *Physical Review E*, 93(2):022126, 2016.

[65] Lucien Hardy. Quantum theory from five reasonable axioms. *arXiv preprint quant-ph/0101012*, 2001.

[66] Lucien Hardy. Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure. *Journal of Physics A: Mathematical and Theoretical*, 40(12):3081, 2007.

[67] Lucien Hardy. Reformulating and reconstructing quantum theory. *arXiv preprint arXiv:1104.2066*, 2011.

[68] Lucien Hardy and William K Wootters. Limited holism and real-vector-space quantum theory. *Foundations of Physics*, 42(3):454–473, 2012.

[69] Philipp Andres Höhn. Toolbox for reconstructing quantum theory from rules on information acquisition. *Quantum*, 1:38, 2017.

[70] Philipp Andres Höhn and Christopher SP Wever. Quantum theory from questions. *Physical Review A*, 95(1):012102, 2017.

[71] Michał Horodecki and Jonathan Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature communications*, 4:2059, 2013.

[72] Vojkan Jakšić and Claude-Alain Pillet. A note on the landauer principle in quantum statistical mechanics. *Journal of Mathematical Physics*, 55(7):075210, 2014.

[73] Peter Janotta, Christian Gogolin, Jonathan Barrett, and Nicolas Brunner. Limits on nonlocal correlations from the structure of the local state space. *New Journal of Physics*, 13(6):063024, 2011.

[74] Peter Janotta and Raymond Lal. Generalized probabilistic theories without the no-restriction hypothesis. *Physical Review A*, 87(5):052131, 2013.

[75] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.

[76] Christopher Jarzynski, HT Quan, and Saar Rahav. Quantum-classical correspondence principle for work distributions. *Physical review X*, 5(3):031038, 2015.

[77] Jian Jiang, Ru Wang, Michel Pezeril, and Qiuping Alexandre Wang. Application of varentropy as a measure of probabilistic uncertainty for complex networks. *Chinese science bulletin*, 56(34):3677–3682, 2011.

[78] Pascual Jordan, John von Neumann, and Eugene P Wigner. On an algebraic generalization of the quantum mechanical formalism. In *The Collected Works of Eugene Paul Wigner*, pages 298–333. Springer, 1993.

[79] Richard Jozsa. Entanglement and quantum computation. *arXiv preprint quant-ph/9707034*, 1997.

[80] Petteri Kaski, Patric RJ Östergård, and RJ Patric. *Classification algorithms for codes and designs*, volume 15. Springer, 2006.

[81] Aleks Kissinger and Sander Uijlen. Picturing indefinite causal structure. *arXiv preprint arXiv:1701.00659*, 2017.

[82] Ioannis Kontoyiannis and Sergio Verdú. Optimal lossless compression: Source varentropy and dispersion. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 1739–1743. IEEE, 2013.

[83] Kamil Korzekwa, Matteo Lostaglio, Jonathan Oppenheim, and David Jennings. The extraction of work from quantum coherence. *New Journal of Physics*, 18(2):023045, 2016.

[84] Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM journal of research and development*, 5(3):183–191, 1961.

[85] Ciarán M Lee and John H Selby. Higher-order interference in extensions of quantum theory. *Foundations of Physics*, 47(1):89–112, 2017.

[86] Ciarán M Lee and John H Selby. A no-go theorem for theories that decohere to quantum mechanics. *Proc. R. Soc. A*, 474(2214):20170732, 2018.

[87] Elliott H Lieb and Derek W Robinson. The finite group velocity of quantum spin systems. In *Statistical Mechanics*, pages 425–431. Springer, 1972.

[88] Elliott H Lieb and Robert Seiringer. *The stability of matter in quantum mechanics*. Cambridge University Press, 2010.

[89] Noah Linden, Sandu Popescu, and Paul Skrzypczyk. The smallest possible heat engines. *arXiv preprint arXiv:1010.6029*, 2010.

[90] Matteo Lostaglio. Quantum fluctuation theorems, contextuality, and work quasiprobabilities. *Physical review letters*, 120(4):040602, 2018.

[91] Matteo Lostaglio, David Jennings, and Terry Rudolph. Description of quantum coherence in thermodynamic processes requires constraints beyond free energy. *Nature communications*, 6:6383, 2015.

[92] Matteo Lostaglio, Kamil Korzekwa, David Jennings, and Terry Rudolph. Quantum coherence, time-translation symmetry, and thermodynamics. *Physical review X*, 5(2):021001, 2015.

[93] Donald Lynden-Bell. Negative specific heat in astronomy, physics and chemistry. *Physica A: Statistical Mechanics and its Applications*, 263(1-4):293–304, 1999.

[94] Lluís Masanes, Yeong-Cherng Liang, and Andrew C Doherty. All bipartite entangled states display some hidden nonlocality. *Physical review letters*, 100(9):090403, 2008.

[95] Lluís Masanes and Markus P Müller. A derivation of quantum theory from physical requirements. *New Journal of Physics*, 13(6):063001, 2011.

[96] Lluís Masanes and Jonathan Oppenheim. A general derivation and quantification of the third law of thermodynamics. *Nature communications*, 8:14538, 2017.

[97] Markus P Müller and Cozmin Ududec. Structure of reversible computation determines the self-duality of quantum theory. *Physical review letters*, 108(13):130401, 2012.

[98] Yûto Murashita, Ken Funo, and Masahito Ueda. Nonequilibrium equalities in absolutely irreversible processes. *Physical Review E*, 90(4):042110, 2014.

[99] Yûto Murashita and Masahito Ueda. Gibbs paradox revisited from the fluctuation theorem with absolute irreversibility. *Physical review letters*, 118(6):060601, 2017.

[100] Miguel Navascués and Harald Wunderlich. A glance beyond the quantum model. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, page rspa20090453. The Royal Society, 2009.

[101] Johann v Neumann. *Mathematische grundlagen der quantenmechanik*, volume 38. Springer-Verlag, 2013.

[102] Ognyan Oreshkov, Fabio Costa, and Časlav Brukner. Quantum correlations with no causal order. *Nature communications*, 3:1092, 2012.

[103] Ognyan Oreshkov and Christina Giarmatzi. Causal and causally separable processes. *New Journal of Physics*, 18(9):093020, 2016.

[104] Congjie Ou, Aziz El Kaabouchi, Laurent Nivanen, Jincan Chen, Franois Tsobnang, Alain Le Méhauté, and Qiuping A Wang. Maximizable informational entropy as a measure of probabilistic uncertainty. *International Journal of Modern Physics B*, 24(17):3461–3468, 2010.

[105] Marcin Pawłowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101, 2009.

[106] Judea Pearl. *Causality*. Cambridge university press, 2009.

[107] JP Pekola, S Suomela, and YM Galperin. Finite-size bath in qubit thermodynamics. *Journal of Low Temperature Physics*, 184(5-6):1015–1029, 2016.

[108] Martí Perarnau-Llobet, Elisa Bäumer, Karen V Hovhannisyan, Marcus Huber, and Antonio Acin. No-go theorem for the characterization of work fluctuations in coherent quantum systems. *Physical review letters*, 118(7):070601, 2017.

[109] Max Planck. On the law of distribution of energy in the normal spectrum. *Annalen der physik*, 4(553):1, 1901.

[110] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[111] David Reeb and Michael M Wolf. An improved landauer principle with finite-size corrections. *New Journal of Physics*, 16(10):103011, 2014.

[112] David Reeb and Michael M Wolf. Tight bound on relative entropy by entropy difference. *IEEE Transactions on Information Theory*, 61(3):1458–1473, 2015.

[113] Jonathan G Richens, Álvaro M Alhambra, and Lluis Masanes. Finite-bath corrections to the second law of thermodynamics. *Physical Review E*, 97(6):062132, 2018.

[114] Jonathan G Richens and Lluis Masanes. Work extraction from quantum systems with bounded fluctuations in work. *Nature communications*, 7:13511, 2016.

[115] Jonathan G Richens, John H Selby, and Sabri W Al-Safi. Entanglement is necessary for emergent classicality in all physical theories. *Physical review letters*, 119(8):080503, 2017.

[116] Carlo Rovelli. Relational quantum mechanics. *International Journal of Theoretical Physics*, 35(8):1637–1678, 1996.

[117] CA Ryan, O Moussa, J Baugh, and R Laflamme. Spin based heat engine: demonstration of multiple rounds of algorithmic cooling. *Physical review letters*, 100(14):140501, 2008.

[118] Sina Salek and Karoline Wiesner. Fluctuations in single-shot ε-deterministic work extraction. *Physical Review A*, 96(5):052114, 2017.

[119] John Selby and Bob Coecke. Leaks: quantum, classical, intermediate and more. *Entropy*, 19(4):174, 2017.

[120] John H Selby, Carlo Maria Scandolo, and Bob Coecke. Reconstructing quantum theory from diagrammatic postulates. *arXiv preprint arXiv:1802.00367*, 2018.

[121] Anthony J Short and Jonathan Barrett. Strong nonlocality: a trade-off between states and measurements. *New Journal of Physics*, 12(3):033034, 2010.

[122] Urbasi Sinha, Christophe Couteau, Zachari Medendorp, Immo Söllner, Raymond Laflamme, Rafael Sorkin, and Gregor Weihs. Testing bornâs rule in quantum mechanics with a triple slit experiment. In *AIP Conference Proceedings*, volume 1101, pages 200–207. AIP, 2009.

[123] Paul Skrzypczyk, Nicolas Brunner, Noah Linden, and Sandu Popescu. The smallest refrigerators can reach maximal efficiency. *Journal of Physics A: Mathematical and Theoretical*, 44(49):492002, 2011.

[124] Paul Skrzypczyk, Anthony J Short, and Sandu Popescu. Work extraction and thermodynamics for individual quantum systems. *Nature communications*, 5:4185, 2014.

[125] Rafael D Sorkin. Quantum mechanics as quantum measure theory. *Modern Physics Letters A*, 9(33):3119–3127, 1994.

[126] Robert W Spekkens. Evidence for the epistemic view of quantum states: A toy theory. *Physical Review A*, 75(3):032110, 2007.

[127] Dan Sperber, David Premack, and Ann James Premack. *Causal cognition: A multidisciplinary debate*. Number Sirsi) i9780198523147. Clarendon Press Oxford, 1995.

[128] Ernst CG Stueckelberg. Quantum theory in real hilbert space. *Helv. Phys. Acta*, 33(727):458, 1960.

[129] Leo Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behavioral science*, 9(4):301–310, 1964.

[130] Peter Talkner and Peter Hänggi. Aspects of quantum work. *Physical Review E*, 93(2):022131, 2016.

[131] Raam Uzdin, Amikam Levy, and Ronnie Kosloff. Equivalence of quantum heat machines, and quantum-thermodynamic signatures. *Physical Review X*, 5(3):031044, 2015.

[132] Frederick Albert Valentine. *Convex sets*, volume 99. McGraw-Hill New York, 1964.

[133] VS Vladimirov and IV Volovich. p-adic quantum mechanics. *Communications in Mathematical Physics*, 123(4):659–676, 1989.

[134] Peter Walters. *An introduction to ergodic theory*, volume 79. Springer Science & Business Media, 2000.

[135] Mischa P Woods, Nelly Ng, and Stephanie Wehner. The maximum efficiency of nano heat engines depends on more than temperature. *arXiv preprint arXiv:1506.02322*, 2015.

[136] William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982.

[137] Wojciech H Zurek. Environment-induced superselection rules. *Physical review D*, 26(8):1862, 1982.

[138] Karol Życzkowski. Quartic quantum theory: an extension of the standard quantum mechanics. *Journal of Physics A: Mathematical and Theoretical*, 41(35):355302, 2008.