



**This electronic thesis or dissertation has been  
downloaded from Explore Bristol Research,  
<http://research-information.bristol.ac.uk>**

*Author:*

**Purves, Tom**

*Title:*

**Quantum Theory**

*Causality, Thermodynamics and Post-Selection*

**General rights**

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

**Take down policy**

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact [collections-metadata@bristol.ac.uk](mailto:collections-metadata@bristol.ac.uk) and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

---

---

# Quantum Theory: Causality, Thermodynamics and Post-Selection

---

---

By

TOM PURVES



Department of Quantum Information  
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Physics.

2021

Word count:76078



## ABSTRACT

This thesis explores the interplay between quantum theory, post selection, thermodynamics and causality. In the first chapter, we review some notation and concepts key to quantum information science. In chapter two, we will present work on channels, measurements and post-selection with an emphasis on the thermodynamics of such objects. In the third chapter, we will expand on what we mean when we say causality is classical in quantum theory. Chapter four goes beyond quantum theory, to a fantasy quantum mechanics that can be simulated via post selection. We show that a local notion of no-backwards-in-time signalling is insufficient to ensure a global causal description of quantum theory. Finally, in chapter five, we will present the resource theory of post-selection.



## DEDICATION AND ACKNOWLEDGEMENTS

This page is the single most important piece contained within this Thesis. It will be impossible to mention everybody, and even more impossible to list the ways each and every person has contributed.

First, I would like to thank and acknowledge my family. My father Graeme, for being my first best friend and for always being my hero. My Mother, Susan for the absolute and unconditional love she has given. My sister Emma, to whom this work is dedicated, for inspiring me every day. To my grandfather Mervyn for all of those encyclopedias gifted and to my Grandmother Janet who will always occupy a special and significant place in my heart.

Next, the academics and scientists who have guided me along the way. Dr. Tony Short, for your friendship, kindness, insight, integrity, and intellect. These gifts are not easily shared but will always be well appreciated. Thank-you to Dr. Paul Skrzypczyk, for successfully explaining the un-understandable always. Patryk and Andreas, for the fun in the office and the friendship. Professor Sandu Popescu, for showing me the meaning of paradox, for always taking my calls, and for making the process of science magical. Thank-you goes to Dr. Jaap Velthuis, for the patience and belief. Finally, a big thank-you to Professor Rennato Renner, for agreeing to act as examiner to this body of work.

I would also like to extend the warmest thanks to coach Saeed Esmali, for growing my character and for gifting me the Pahlevan. I would also like to take the opportunity to extend my support for the wrestling for humanity initiative and all of the fantastic work the social project has already and will go on to achieve. All of my teammates are also not without thanks and gratitude.

Jack, Tom, Ben - thanks lads, dilly dilly.

For steering this ship with me for the past four years through storms and fair weather I must thank Ben, Oscar, Nick, Liam, Abbey, Jenna, Kira, Andy, Crabby, Dan, Luke, Luca, Ling, Sammy, George, Rachel, and Prasan. Best wishes for the future are gratuitously extended.

Sasha, thank-you, from the very bottom of my heart.



## AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: THOMAS PURVES

DATE: DECEMBER 12, 2022





## TABLE OF CONTENTS

	Page
<b>List of Tables</b>	<b>ix</b>
<b>List of Figures</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Quantum Theory . . . . .	2
1.1.1 History and Background . . . . .	2
1.1.2 Postulates . . . . .	4
1.1.3 Formalism . . . . .	8
1.1.4 Thermodynamics . . . . .	13
1.1.5 Bell Inequalities . . . . .	16
1.1.6 The Problem of Defining Time . . . . .	20
1.2 Causality . . . . .	25
1.2.1 History . . . . .	25
1.2.2 Formalism . . . . .	27
1.2.3 Levels of the Pearl Hierarchy . . . . .	31
1.2.4 Quantum Causal Models . . . . .	33
1.2.5 Causal Inequalities . . . . .	36
1.3 Post-Selection . . . . .	40
1.3.1 Connections to Quantum Mechanics . . . . .	40
1.3.2 Self-Controlled Unitary Gates . . . . .	44
1.3.3 Complexity Theory . . . . .	46
1.3.4 Connections to Causality . . . . .	47
1.4 This Work . . . . .	47
<b>2 Channels, Measurements and Post-Selection in Quantum Thermodynamics</b>	<b>49</b>
2.1 Introduction . . . . .	49
2.2 Framework . . . . .	50
2.2.1 Implicit Batteries . . . . .	51
2.2.2 Explicit Batteries . . . . .	55

TABLE OF CONTENTS

---

2.3	Channels . . . . .	60
2.4	Measurements . . . . .	68
2.5	Post-Selection . . . . .	80
2.6	Discussions and Conclusions . . . . .	83
<b>3</b>	<b>Causality is Classical in Quantum Theory</b>	<b>85</b>
3.1	Introduction . . . . .	85
3.2	Causal Process . . . . .	87
3.3	Quantum Process . . . . .	91
3.4	Main Result . . . . .	93
3.5	Quantum Switch . . . . .	103
3.6	Equivalence Between Combined And Individual Controlled Lab Gates . . . . .	107
3.7	Discussions and Conclusions . . . . .	109
<b>4</b>	<b>Non-Classically Causal Correlations Without Backwards-In-Time Signalling</b>	<b>111</b>
4.1	Introduction . . . . .	111
4.2	Two-Time States . . . . .	113
4.3	The NBTS Paradigm . . . . .	119
4.4	Linear Two-Time States . . . . .	122
4.5	Procedure . . . . .	125
4.5.1	Methods For Generating The Data . . . . .	132
4.6	Discussions and Conclusions . . . . .	134
<b>5</b>	<b>The Resource Theory of Post-Selection</b>	<b>137</b>
5.1	Introduction . . . . .	137
5.2	Resource Theories . . . . .	139
5.3	Quantum Resource Theory Example . . . . .	143
5.4	Resource Theory of Post-Selection . . . . .	144
5.5	Discussion and Conclusion . . . . .	155
<b>6</b>	<b>Conclusions</b>	<b>159</b>
6.1	Discussions from Thermodynamics . . . . .	159
6.1.1	Further Works . . . . .	159
6.2	Discussions from Causality . . . . .	160
6.2.1	Further Works . . . . .	160
6.3	Discussions from Post-Selection . . . . .	161
6.3.1	Further works . . . . .	161
6.4	Outro . . . . .	161
	<b>Bibliography</b>	<b>163</b>

## LIST OF TABLES

<b>TABLE</b>	<b>Page</b>
4.1 Table of probabilities generated by the protocol. . . . .	130



## LIST OF FIGURES

FIGURE	Page
1.1 The Bloch sphere representation of qubits. Pure states live on the surface, mixed states live in the interior of the Bloch sphere. Pure states are parametrised by their angle to the $\sigma_z$ axis and $\sigma_x$ axis. . . . .	11
1.2 The hidden variable common cause explanation for the Bell scenario. There exist experiments in quantum theory that cannot be explained in this manner. . . . .	17
1.3 A 2D slice through the Local and No-signalling polytopes. The set of quantum achievable correlations $\mathcal{Q}$ sits between these two polytopes. PR boxes sit at the vertices of the No-signalling set. Local is the set of all correlations that can be written with a classical common cause model. . . . .	19
1.4 Before the scattering, the particles are in quadrant one, and travelling towards the negative $x$ and the negative $y$ axis. Afterwards, classically, the particles will be in either quadrant three or four. Quantum mechanically there are solutions where the particles are found back in quadrant one, or even in quadrant two! . . . . .	23
1.5 A simple causal diagram relating the air pressure, a reading on a barometer needle, and whether the weather is a thunderstorm. . . . .	32
1.6 The two permissible DAGs for causal models in the causal scenario described. These cover all possible probabilities consistent with the temporal ordering $t_x < t_a < t_y < t_b$ on the left and $t_y < t_b < t_x < t_a$ on the right, with independent inputs. If a bipartite probability distribution cannot be explained as arising from a mixture of these causal models, and there are no latent variables in the model (i.e. the system is well isolated) then we say that the correlation is not causal, and violates a causal inequality. All causal correlations for two parties can be written as in equation (1.57). . . . .	38
1.7 Circuit diagram notation for a controlled gate. . . . .	44
1.8 Circuit diagram notation for a self-controlled gate. . . . .	44
1.9 Cap-and-cup diagram of the process used to build a self controlled gate utilising post-selection and a standard controlled unitary gate. The bras on the right hand side indicate a successful post-selection upon the state $\frac{ 00\rangle+ 11\rangle}{\sqrt{2}}$ . . . . .	46

LIST OF FIGURES

---

2.1	Schematic description of a thermal operation in the implicit battery framework. The work is an exogenous quantity which ensures the first law of thermodynamics holds.	52
2.2	Schematic description of the thermal operations in the explicit battery framework. The work is an endogenous property of the system, and is represented by the work done to raise or lower a weight system. This energy can be used at a later time for some other process.	55
2.3	Schematic description of the protocol $P$ , which consists of a channel application and a unitary reset. A single use of a quantum channel is used to provide a work advantage in a cyclic protocol. As in the main text, the system is composed of an ancilla and target $S = A \otimes T$ , the bath is represented by $B$ and the weight system by $W$ .	62
2.4	Illustration of the dilation theorem, where a quantum channel is realised by a unitary operation on a larger system.	66
2.5	Illustration of a single cycle of a thermodynamic protocol $P$ , which consists of a measurement and a unitary reset. Different unitary resets are applied, dependant upon the measurement outcome obtained in the previous step, to give some feed-forward control over the extractable work.	71
2.6	Three possible weight states. In a) we depict the top hat weight state, in b) we depict a real symmetric weight state. In c) we depict a weight state that is real, but not symmetric.	78
3.1	Simple representation of Alice's local laboratory, during an experiment. At some point, a system enters her lab. A measurement is made of that system which returns outcome $\alpha$ , conditional on some variable $x$ . The system is then permitted to exit her lab.	87
3.2	A simple causal process, where Alice is first, Bob is second and Charlie is third in the causal order. This process is deterministic, in the sense that the causal order is not associated to some probability which is $< 1$ . What goes on inside the labs however, might well be probabilistic.	88
3.3	Two possible causal scenarios.	90
3.4	Quantum circuit description of the generic measurement procedure for quantum theory. On the left-hand side, we have the initialisation step, where the Control, System, Results and Flag registers are initialised to the zero state. A sequence of $\hat{U}$ and $\hat{V}$ operators are then applied in order to measure the system. The measurements are achieved by an entangling unitary between the system, results and flag spaces for each party. At the end of the protocol, we demand that the flags are all raised, and then a projective measurement is performed on the results registers. By repeating this procedure many times, $p^{quantum}(\vec{a} \vec{x})$ are constructed.	93
3.5	A schematic of where the states $\phi$ and $\psi$ appearing in definitions 3.3 and 3.2 live in the protocol.	94

---

3.6	Cartoon of what the result says; on the left hand side we have a causally indefinite process, on the right hand side, we have its simulation with a probabilistic causal model.	104
3.7	Realisation of the quantum switch in our framework through a quantum circuit. . . .	105
3.8	A controlled lab gate for an individual laboratory . . . . .	107
3.9	An equivalent circuit to the individual controlled lab gate above, built from a single instance of $V$ , where $W_l 0\rangle =  l\rangle$ . Note that the individual wires may represent composite subsystems rather than individual qubits. . . . .	107
3.10	$V$ , built from individual controlled lab gates. The first and last $CNOT$ gate are controlled from the state $ 1\rangle_c$ , the second and second from last are controlled from $ 2\rangle_c$ , and so on until the $n$ 'th and $n + 1$ 'th $CNOT$ , which are controlled from state $ n\rangle_c$ . . .	108
4.1	A slider puzzle, where the rules are that the blocks can only be moved in direction of their long axis, and the walls of the $5 \times 5$ grid are hard boundaries. Here is the solution: "A7, 2U; A6, 1L; A2, 2D; A4, 1R". And here is a jumbled causal order solution: "A2, 2D; A4, 1R; A6, 1L; A7, 2U". The jumbled causal order solution does not solve the problem (and actually, for this slider problem, there is only one move possible at each time-step, so there is no advantage at all in the jumbled order solution). The order in which you follow the steps in the solution is clearly a resource in the problem. . . . .	113
4.2	A diagram of a quantum process involving post-selection. At some early time $t = t_1$ , the system is prepared in a state $ \phi\rangle$ . An experiment is conducted and, at a later time $t = t_2$ , the system is post-selected into the state $ \psi\rangle$ . . . . .	115
4.3	A two-time state experiment. . . . .	117
4.4	Two experiments, equivalently described by standard quantum theory and post-selection or a two-time state. . . . .	118
4.5	Depiction of the experiment described by equation (4.13). . . . .	119
4.6	Pictorial representation of a laboratory in the NBTS paradigm. At some point, a system enters into a laboratory. A measurement of that system is made, and then the system gets transformed (based on some random variable $x$ ). The system then leaves the lab. . . . .	120
4.7	A state for a single party that ensures the NBTS conditions are met, irrespective of the measurements that are made in the middle. This corresponds to a preparation of a quantum state, and a trivial post-selection. . . . .	123
4.8	A picture of the protocol that achieves the NBTS, but causal inequality violating statistics described in the main text. The black lines connecting laboratories represent channels with a $z$ -basis measurement. The gates labelled $X$ are flip gates. Notice that in the local frame of each laboratory the output is produced before the input is received.	127





## INTRODUCTION

Quantum theory is a set of fantastic tools for computing quantities about the physical world in which you and I live. Indeed, it is often touted that quantum theory is ‘the best physical theory we have’. Yet the world of these quanta, the world that ‘they’ live, is vastly different from our own. Attempts to reason about this quantum world are muddled, messy, noisy and often, and perhaps worst of all, paradoxical [1]. How can it be, that by knowing things about all the constituent parts of a system we know nothing of the whole. Or how can it be that cats exist as alive feeling creatures and simultaneously exist as dead clumps of matter. How we explain quantum phenomena has been greatly strengthened in recent years through information theory. Although a warning, information in the quantum realm is of a vastly different character than in the classical case. Is it our fundamental inability to reason about these quantum types of information why we struggle so much to understand quantum mechanics, and why it is a common topic of discourse among serious philosophers of science today how to interpret quantum mechanics? This thesis will make attempts to contribute to the answers to this field. Here, we will reason about and focus on physical principles, mostly the principles of locality and causality, to see if we can hold mirrors up against the quantum realm, and see structures in that world which replicate shapes in our own. I hope that you will enjoy the discourse, and along the way discover that although the quantum world is certainly strange, it is also beautifully constructed, with glimpses of our own below the surface.

## 1.1 Quantum Theory

### 1.1.1 History and Background

Let me for a short few words begin by setting the stage with a class in the history of quantum theory. Every student of quantum mechanics, be they now an expert in quantum computing or quantum encryption or quantum gravity begins this way, and for good reason - one must be shocked. The first shocking idea we will arrive at, is that of a quanta. Imagine you have a pipe, and the pipe is full of an extremely slippery substance, say oil. We are interested in transporting oil from one location to another, and the obvious thing to do is to lay the pipe between the two locations and pump oil from one end of the pipe. Because oil is not a compressible substance, and the pipe is already full, any oil that you pump into the pipe from side A will fall from the pipe at side B (we are assuming here that the cross section of the pipe is constant). Moreover, in the ideal case, the rate at which you pump oil from the side A will match the rate oil leaves the pipe at side B. This in some way represents a transfer of *energy* from one location to another. In the classical world, we may make the pump rate at side A as low as we like, there is no minimum energy required to begin the process of the oil falling from side B. In the quantum mechanical case, not so! The photoelectric effect [2], the first experiment that hints at a quantum mechanical description can be thought of as similar to our oil pipe. In the photoelectric effect, light is shone onto the surface of a metal and the current caused by the incident radiation observed. What one finds is that, once we use the kinetic energy of electrons as a proxy for the current, it follows a simple rule

$$(1.1) \quad E_{max} = hf - \phi$$

where  $\phi$  is a quantity known as the work-function and intrinsic to the type of metal we use in the experiment,  $f$  is the frequency of incident light and  $h$  is the Planck constant, which takes a value of  $h = 6.626 \times 10^{-34} \text{ JHz}^{-1}$ . The striking thing is that since kinetic energy is strictly a positive quantity, there is a minimum frequency of light that one must shine to observe any current at all. You can do this experiment at home with a surface of metal, a soldering iron and a source of monochromatic light. We recall that frequency and energy of light are related through the famous Einstein formula

$$(1.2) \quad E = hf.$$

This is remarkable. This is also where the development of the quantum theory begins. The Schrödinger equation, the Heisenberg uncertainty principle, and Feynman diagrams all come after and are children of this remarkably simple observation - that for electrons and light, unlike oil in pipes, there is a minimum amount of energy one must press into the pipe in order to begin the flow. It is almost like our pipe of oil is no longer a continuous medium of slippery fluid, but neatly arranged into little buckets that only move along the pipe when the bucket at side A has

been filled. One may think of a Japanese water garden, where the clack of a tipping bucket is only heard once the bucket has become sufficiently filled to the tipping point. This minimum energy is called a quanta, and how the discrete emerges from the continuous was the focus of the pioneers of quantum theory in the early 1900's.

It is a fact of electromagnetism that an accelerating charge radiates energy [3]. The derivation of this fact is long and cumbersome, but involves only Maxwell's equations and some calculus and this fact soon stares us in the face. How then, is a negatively charged electron able to orbit the centre of an atom (also called the nucleus, which we know today to be made of positively charged protons and neutral neutrons)? Orbit is a feat of constant acceleration, which is towards the focus. Where is the radiated away energy, and moreover, if an electron is constantly radiating energy away, should it not fall ever closer to the centre of the atom, spiralling further to its eventual doom in collision with the atom centre? This doom, by the way, would soon end any physical structures in the universe including you and I, for without atoms there would be no molecules, no planet to stand on and no DNA to form life. The resolution to this paradox is quantum theory, and we have since learned that not only energy, but other physical quantities, such as angular momentum can and are quantized into quanta. This calculation was performed by Neils Bohr in 1913 [4] and interpreted by De Broglie in 1924 [5] as a standing wave condition on the allowed stable orbits of electrons around the centre of atoms. Bohr found that the stable radii and energies of the electrons orbiting the atom are

$$(1.3) \quad r_n = \frac{\epsilon_0 h^2}{\pi m_e e^2} n^2$$

and

$$(1.4) \quad E_n = -\frac{m_e e^4}{8\epsilon_0^2 h^2} \frac{1}{n^2}$$

and with this model went on to explain mysterious effects of gases where spectral lines appeared at discrete intervals, which we now know are caused by electrons hopping between discretely defined orbital energies, stimulated by some incident light of sufficient energy.

The interpretation of a particle, such as an electron, to have wave like properties such as standing nodes is no accident, or happy coincidence. One of the lessons learned by the study of quantum theory is that there is no absolute distinction between particles and waves. They are different ways of looking at, or to say - manifestations - of the same underlying physical object, be it an electron or a photon or a proton.

Guiding the forefathers of quantum mechanics in their study of quanta were physical principles. An example of a physical principle is the conservation of energy. Without this, one needs not to appeal to quanta to explain the photoelectric effect. We do not need to analogise about little buckets or anything of the sort, but we also cannot therefore explain the photoelectric effect in terms of anything that we know to be true - and so cannot answer the question of 'why?'. Another example of a physical principle is the conservation of momentum. Another example is

the conservation of angular momentum. These three principles can be derived from both classical mechanics in its Hamiltonian or Lagrangian formalism by making use of Emmy Noethers theorem relating symmetries and conservation laws [6]. However, a physical principle does not have to be a conservation law, there are more kinds of physical principles which we will meet along this discourse which may prove useful in guiding our insights into the world of quantum theory. It seems to be a tenet of quantum theory that it is the theory that is as strange as it can possibly be, while respecting physical principles. To explain it the best one can do is to appeal to those physical principles. The reader may draw analogies between physical principles and axioms from pure mathematical study, and at this point I would encourage them to stop. Axioms form a bedrock from-which a mathematician can build any kind of theory he wishes about anything. A physical principle, to a physicist is only valid in so far as there are no experiments which contradict it. We must always be ready to sacrifice our theories based on physical principles should the physical principle be shown to be false. The mathematician has the luxury of being able to define axioms however he so desires, it is then down to him to make sense of the theory. Physics, has to respect natural law. Hence, we often hear physicists talk of ‘regimes of validity’, or ‘approximations holding when’, statements not of the weakness of his theory but of the physical principles he has applied when coming up with such a theory. For instance, the trajectory of a ball in air may be calculated using the physical principle of energy conservation, however at high wind speeds the trajectory observed may differ with the theory - not because his theory was wrong, but because the physical principle of conservation of energy applied between the ball and the earth has neglected to take into account the energy carried in the wind.

### 1.1.2 Postulates

In courses in quantum theory students are presented with about five or six postulates that quantum theory adheres to. While these vary from author to author, dependent of course on regions of interest and viewpoint of the academic stating them, in this section I will lay out my favourite so that for the remainder of the thesis one will know precisely what I intend to mean when I say quantum theory. All of quantum mechanics then follows from these assumptions which cannot themselves be derived (see next section for an explanation of the notation used in these postulates).

**Postulate 1.** *At each instance the state of a physical system is represented by a vector in the Hilbert space of states.*

While this postulate seems benign philosophical jargon, it is anything but. This postulate already makes the statement that superposition occurs in quantum theory. If we have that  $|\phi_1\rangle$  and  $|\phi_2\rangle$  are possible states for the system, then we will also find that  $|\phi\rangle = a_1|\phi_1\rangle + a_2|\phi_2\rangle$  a valid state of the system, from the structure of Hilbert space itself. In the example of Schrödinger’s cat, since  $|\text{cat alive}\rangle$  and  $|\text{cat dead}\rangle$  are indeed valid states of a cat, then so is the superposition

$|\psi\rangle = \alpha_{alive} |\text{cat alive}\rangle + \beta_{dead} |\text{cat dead}\rangle$ . Do not be concerned, the remaining postulates are just as strange. It is also worth noting at this point that the space of states, being a Hilbert space, come equipped with the notion of inner product, which is a sort of distance or overlap measure, and in quantum theory takes the form

$$(1.5) \quad \langle a|b\rangle = \int dx a^*(x)b(x).$$

With this inner product comes the definition of the dual space to the set of states, and the star denoting complex conjugate. If the underlying Hilbert space is discrete, this inner product can instead be written using the ‘natural isomorphism’ between integrals and summations,

$$(1.6) \quad \langle a|b\rangle = \sum_{i=0}^d a_i^* b_i.$$

Dual spaces are of immense importance in quantum mechanics. In Hilbert space there is a natural isomorphism existing between the dual and primal vector spaces, in the form of the dagger operator. The dagger operator represents the complex conjugate transpose, i.e.  $a^\dagger = (a^*)^T$ . We will be often assigning the curly H ‘ $\mathcal{H}$ ’ to Hilbert spaces. Finally, the wording ‘at each instance’ is not meant to trip the reader into a false sense of security, and it means the same thing as an event does in relativity.

**Postulate 2.** *Every quantity that can be learned from a state is an observable, associated to a Hermitian operator that acts on the Hilbert space of states.*

A Hermitian operator is an operator  $A$  which carries the property that  $A = A^\dagger$ . Hermitian operators carry some nice properties. Their eigenvalues are real, which can be proven in just a few lines;

$$(1.7) \quad a \langle a|a\rangle = \langle a|A|a\rangle = \langle a|A^\dagger|a\rangle = (\langle a|A|a\rangle)^\dagger = (a \langle a|a\rangle)^\dagger = a^* \langle a|a\rangle$$

and their eigenfunctions are orthogonal to one another, again a few lines; suppose that  $A|a\rangle = a|a\rangle$  and  $A|a'\rangle = a'|a'\rangle$ . We know the scalars in these equations are real, and so we have  $\langle a'|A|a\rangle = a \langle a'|a\rangle$ . We also have by using the defining property of being Hermitian, that  $\langle a'|A|a\rangle = a' \langle a'|a\rangle$ , and by combining results we find that  $(a - a') \langle a'|a\rangle = 0$ , from which orthogonality follows. The niceness of Hermitian operators does not end there. Given any Hermitian operator we can find a set of eigenstates that will form an orthogonal basis to our Hilbert space. This basis is often called an eigenbasis and we now see why this part of the postulate of quantum mechanics is essential. It assures us that the types of questions we can ask of quantum states at least have a set of sensible answers. This is followed and best surmised by the next postulate

**Postulate 3.** *The only results of measurement of a Hermitian observable  $A$  are the eigenvalues of the operator  $\hat{A}$*

This postulate ensures a few important things. Firstly, when we ask a question of a state in quantum mechanics, it will return us a real-valued answer. This is important since without this restriction one may imagine, given the plethora of complex-valued quantities in quantum mechanics, we may if not careful, get silly answers such as ‘the position of the particle is  $i$ ’ with  $i$  being the complex unit  $i = \sqrt{-1}$ . Secondly, note here that the answer that you get in quantum mechanics depends on the question which you ask. A quantum state can describe many things simultaneously, but if you wish to know something about the position of a particle described by that quantum state, you had better make sure to ask a question which pertains to position! In this example, this would correspond to some kind of manipulation involving both the quantum state  $|\phi_x\rangle$  and the position operator  $\hat{x} = \int dx x|x\rangle\langle x|$ . We discuss what that manipulation is in the next section. In other words, how do I use quantum theory to answer questions?

**Postulate 4.** *The probability to observe outcome  $a$ , an eigenvalue of a Hermitian observable  $A$ <sup>1</sup>, given that the state of the system is well described by a wave function  $\phi$  is given by*

$$(1.8) \quad p(a) = |\langle a|\phi\rangle|^2,$$

where  $|a\rangle$  is the eigenstate associated to  $a$ , and the operation  $|\cdot|^2$  is the complex norm. Immediately after the measurement, the quantum state of the system is in the eigenstate  $|a\rangle$

This postulate departs from classical theories of nature with the introduction of probability. Quantum mechanics is a fundamentally *probabilistic* theory of nature. This is perhaps the biggest way that classical and quantum physics depart from one another, the source of many paradoxes in quantum theory. Newtons mechanics is a deterministic theory, if total knowledge of the forces and masses are present then one can in principle solve a system for all time, and moreover be sure that any experiment is proposed to test Newtonian mechanics, then the answers to any questions we ask of that experiment will be the same for each run of the experiment, up-to experimental errors. Special and general relativity also have this property, they follow the pattern of initial conditions plus equations of motion give deterministic answers for the evaluations of systems for all time. In classical physics, probability was devised to compensate for a lack of knowledge about a system. A good analogy is that of rolling dice. If we arrange a robot who has perfect control over his limbs and an extremely fancy trajectory analysis software running in his operating system, it is in principle possible for this robot to roll snake eyes with a pair of dice every time he decides to roll. The inability for humans to determine with such fine magnitude the initial orientation and forces needed to do so is why casinos feel safe to still play dice games for large sums of money. It is only the human lack of knowledge about how to throw the die that makes his outcomes appear probabilistic. Sometimes I get a six, sometimes I get a one and sometimes the numbers in between. Quantum theory is not like this classical world, and there is no way around using a probabilistic description for the answers to questions which we ask about a quantum state. A

---

<sup>1</sup>with a distinct set of eigenvalues and eigenvectors

quantum state already encodes all the information that we have to hand about a quantum system, so the cause of the probabilities does not reside in some ignorance. It is instead fundamental to the wave function, and no classical robot can predict *ex ante* the results of throwing quantum dice.

One corollary often quoted along side this postulate is that it is imperative to normalise quantum states to unity, which is to say divide the state by a number such that  $\langle\phi|\phi\rangle = 1$ . This ensures in turn that the probability for any answer to a quantum question is a number that ranges from 0 to 1. For vanilla quantum theory we too will make this assignment, however in the further chapters of 3 and 4 we will depart from this convention in order to make sense of a post-selected quantum theory. When I am to do so, I will endeavour to make this clear. We will also ignore the case of operators with degenerate eigenvalues (operators where the eigenvalues are common among distinct eigenvectors, a property known as having a geometric multiplicity above 1).

The last juxtaposition of physics and mathematics carried with this postulate is the famed ‘collapse’ of the wave function. The collapse of the wave function is an essential ingredient for the Copenhagen interpretation of quantum mechanics, however some competing interpretations, many worlds or the pilot wave theory for example, do not require such a collapse. Whether a physical effect or simply an epistemic effect (which is to say concerned with our state of knowledge about a quantum system) it is true that should another measurement of the same observable be made instantaneously after the first, we can be sure that the second measurement will return the same answer as the first. Simply this is because we are measuring the Hermitian observable  $A$  on the state  $|a\rangle$ , from which the statement follows.

Certainly for classical physics, there is no fundamental reason why measuring one quantity  $A$  of a system will interfere with subsequent measurements of other quantities  $B$ <sup>2</sup>. I am permitted to measure both positions and momenta in classical physics and the order in which I do this doesn’t matter. In quantum theory, this is only true if the state of the system is a simultaneous eigenstate of both  $A$  and  $B$ . For commuting quantities this can and does happen, however for the position and momentum example just mentioned, this is forbidden. In general, for all but well defined specific examples, measuring  $A$  first does change the value you observe by measuring  $B$  (or at least, the distribution of values you measure in  $B$ ). The issue of measurement order will be imperative to the understanding of this thesis in its entirety, but there is nothing in it that this postulate does not completely describe.

**Postulate 5.** *The time evolution of a quantum state is given by the Schrödinger equation.*

This postulate is concerned with how quantum states of a system evolve in time. The Schrödinger equation itself is given by

$$(1.9) \quad i\hbar \frac{d}{dt} |\phi\rangle = \hat{H} |\phi\rangle$$

<sup>2</sup>This is not true in general experimentally, even for the best measurement apparatus, but it is possible in principle



where  $\hat{H}$  is the Hamiltonian of the system. The Hamiltonian is a Hermitian operator and its corresponding observable is the energy of the system. This differential equation is our guide to telling us how a given system will evolve. Our instants from postulate one have been connected in time. One of the most fascinating aspects of quantum theory is that while the process governing evolution of a system in time is deterministic, the measurement process is stochastic and based on a probabilistic interpretations. Many efforts have been made to reconcile these pictures of quantum theory.

Here-on in, we will work in units where  $\hbar = 1$ .

**Postulate 6.** *Given a quantum system in a Hilbert space  $\mathcal{H}_a$  and a second in  $\mathcal{H}_b$ , the resultant system is well described by a state in  $\mathcal{H}_a \otimes \mathcal{H}_b$ , where  $\otimes$  denotes the tensor product.*

This postulate tells us how to combine the descriptions of small quantum systems into a big one, and is also where the other famous quantum effect, entanglement, enters. We will not justify this postulate physically however it is important to think about why this should be the case, and what alternatives might there be. For example, according to this postulate, a two particle wavefunction in position space can be described by  $|\phi_{1,2}\rangle = |\phi_1\rangle \otimes |\phi_2\rangle$ . The normalisation condition for this ensemble, according to postulate 4 is  $\int dr_1 dr_2 |\phi_{1,2}(r_1, r_2)|^2 = 1$ . However, why not describe the entire system with a state  $|\phi\rangle$  such that  $\int dr |\phi|^2 = 2$ ?

It is also of note that the number of fundamental postulates for quantum mechanics have been decreasing. I mentioned that there are a variety of ways to phrase quantum theory. The phrasing is author dependant. One can argue that the fewer the postulates, the better. Einsteins general relativity for example, is a very beautiful theory. It only has two postulates. Recent efforts in the foundations of quantum mechanics theory, initially with Lucian Hardy's Five reasonable axioms for quantum theory [7], then with subsequent works have reduced the number of postulates needed from five right down to three [8]. Although here, in this work, brevity is not the essence of beauty - as so we state quantum theory as you might find it in a famous textbook on the subject [9].

### 1.1.3 Formalism

In this section we will describe the formalism which we will commonly appeal to in order to write down quantum theory. Central to quantum theory is the idea of state and for expressing states we will appeal to Dirac notation. When talking to mathematicians, a vector is any element of a vector space. However, in physics a vector is often reserved for quantities such as position or velocity which appeal to the three dimensions of physical space. As such, it is necessary to make the special distinction for elements of the complex Hilbert space to which quantum states belong. We begin by defining a Hilbert space.

**Definition 1.1.** A Hilbert space  $\mathcal{H}$  is a complex inner product space that is also a complete metric space with respect to the distance function induced by the inner product [10]. A complex inner product space can be surmised by the three following properties

1.  $\langle x|y\rangle = \langle y|\bar{x}\rangle$  for all  $x, y \in \mathcal{H}$
2.  $\langle ax + by|z\rangle = a\langle x|z\rangle + b\langle y|z\rangle$  for all  $x, y, z \in \mathcal{H}$
3.  $\langle x|x\rangle > 0$  for all  $x \neq 0 \in \mathcal{H}$

and the completeness property entails considering that a series of vectors  $x_n$  converges absolutely, which is to say that if  $\sum_{n=0}^{\infty} |x_n| < \infty$  then the series converges in  $\mathcal{H}$ , in the sense that all partial sums converge to an element of  $\mathcal{H}$ .

The significance of Hilbert spaces to quantum theory cannot be overstated. The mathematically rigorous formulation of quantum theory is attributed to John Von-Neumann. Von-Neumann was a polymath, and his influence is stretched over the foundations of computing, mathematics and information theory, however the importance of Hilbert spaces to quantum theory was certainly known prior to the developments of Von-Neumann. The use of the Dirac bra-ket notation is then the following assignment of elements in a Hilbert space.

**Definition 1.2.** A ket vector is an element of a complex valued Hermitian Hilbert space,  $|x\rangle \iff x \in \mathcal{H}$ . A bra is a linear functional living in the dual space  $\mathcal{H}$ , often denoted  $\mathcal{H}^*$  or thanks to Hermiticity  $\mathcal{H}^\dagger$ . In other words we have that  $\langle x| : \mathcal{H} \rightarrow \mathbb{C}$ . We can then combine bras and kets through the inner product. It is also often taken that  $|x\rangle^\dagger = \langle x|$  although this is something of a shorthand since the ket vector is a vector in the Hilbert space and the bra is a linear functional on vectors in  $\mathcal{H}$ , in other words the bra  $\langle x|$  is a combination of a vector and an inner product. Note that what is written inside a ket vector has no physical meaning at all, the state  $|?\rangle$  from the example of Schrödinger's cat is just as valid as the state  $|x\rangle$ .

The distinction between bras and kets is something we will meet again once we begin to describe the two-time formalism of quantum theory. Hilbert spaces of quantum theory permit a variety of tools to be used to solve problems in quantum mechanics. To name a few we are permitted to use the well loved tools of partial differential equations to describe changes in systems, but also and not less importantly we are permitted to use spectral theory. This amounts to taking spectral decomposition's of the operators acting on Hilbert spaces, Sturm-Loiville theory which permits the usage of Greens functions to solve problems pertaining to time evolution and Fourier analysis, wherein we discover that operators naturally exist as Fourier transform pairs with canonical commutation relations that lead to important physical principles, such as the Heisenberg uncertainty relation.

For finite dimensional Hilbert spaces, it is common to express the inner product as a matrix multiplication of a row vector and a column vector

$$(1.10) \quad \langle x|y\rangle = x_1^* y_1 + \dots + x_n^* y_n = \begin{bmatrix} x_1^* & \dots & x_n^* \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ x_m \end{bmatrix}$$

so that the ket vector is identified with

$$(1.11) \quad |y\rangle = \begin{bmatrix} y_1 \\ \vdots \\ x_m \end{bmatrix}$$

and the bra with

$$(1.12) \quad \langle x| = \begin{bmatrix} x_1^* & \dots & x_n^* \end{bmatrix}$$

In quantum theory it is common place to talk of kets with an infinite norm, and so these vectors by completeness do not belong to the Hilbert space they claim to! Examples include kets that represent plane waves of infinite extent or Dirac delta distributions. The mantra here is that we can bring a state arbitrarily close to one of these states and still be within the Hilbert space, but for the convenience of calculation we simply ignore the annoying fact that there is no way to normalise a plane wave in quantum theory. Such issues have been discussed at length by many authors and will be glossed over in this thesis.

We also have to address the issue of state and global phase. It is a fact that can be seen from postulate 4 of quantum theory the states  $|x\rangle$  and  $|x'\rangle = e^{-i\theta} |x\rangle$  are indistinguishable under any question we can ask of the quantum state. As such, we define the physical state to be the equivalence class of all states (also called rays) which differ only in some global phase factor. While global phases are not observable in quantum theory, relative phases certainly are and are the cause of many interesting interference patterns in quantum scattering experiments.

With this all in mind, we can now move to the discussion of the concept of a density matrix. We have just discussed that a quantum state is described by a state vector of a system, denoted  $|\psi\rangle$ . However, when calculating quantum mechanical probabilities we often are only interested, or worse, only have access, to part of a Hilbert space. In this paradigm [11],

1. States are not rays.
2. Measurements are not orthogonal projections.
3. Evolution is not unitary.

We will see that having access to only a part of Hilbert space one finds it becomes necessary to consider classical ensembles of quantum states. So lets do that, and take the case where it is also possible for the quantum state to be in a statistical ensemble of states  $|\psi_i\rangle$ , each state occurring with some classical probability  $p_i$ . The density matrix for the system is then defined as follows.

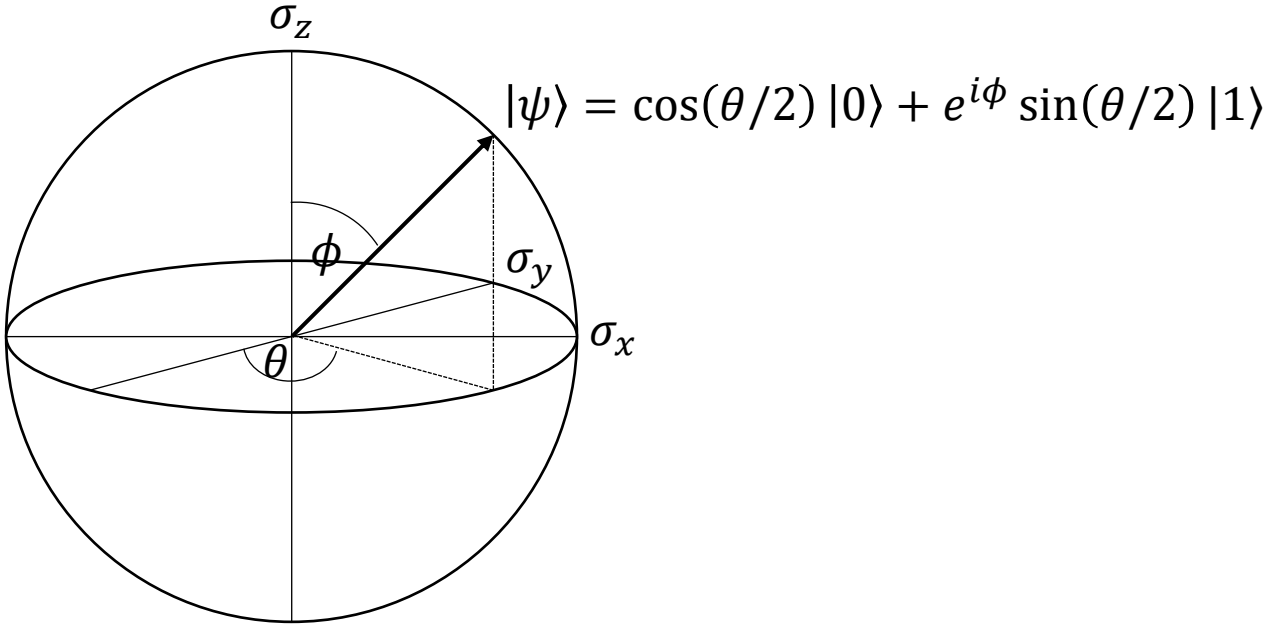


Figure 1.1: The Bloch sphere representation of qubits. Pure states live on the surface, mixed states live in the interior of the Bloch sphere. Pure states are parametrised by their angle to the  $\sigma_z$  axis and  $\sigma_x$  axis.

**Definition 1.3.** *The density matrix of a statistical ensemble of states  $|\phi_i\rangle$  each occurring with probability  $p_i$  is given by*

$$(1.13) \quad \rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|$$

where the operation  $|\psi_i\rangle \langle \psi_i|$  denotes the outer product between the Hilbert space  $\mathcal{H}$  and its dual. The density matrix is positive semi-definite Hermitian operator with trace 1 acting on the state space.

when we have that  $p_i = 1$  for a single  $i$  and all other probabilities are 0, we have a density matrix that describes a *pure* state. We can also take as a definition of a pure state the property that a pure density matrix squares to itself  $\rho^2 = \rho$ , a property known as being idempotent. Any density matrix which is not idempotent is called mixed. A nice geometrical interpretation for the state space exists since all density matrices are convex linear combinations of pure states. For qubits, which are two level quantum systems, this representation is known as the Bloch sphere. It is also important that the representation of a density matrix may not be unique, by which we mean that there may be different ensembles of pure states that lead to the same density matrix.

We are now in a position to discuss how density matrices evolve. We have from the fifth postulate that evolution of the states is given by the Schrödinger equation. Solving the Schrödinger equation for states with the initial condition  $|\psi(t_0)\rangle$  leads us to the equation

$$(1.14) \quad |\psi(t_1)\rangle = U(t_1, t_0) |\psi(t_0)\rangle$$

where the operator  $U(t_1, t_0)$  is a unitary operator. A unitary operator is any operator, in our case on the Hilbert space  $\mathcal{H}$ , that satisfies the relation  $U(t_1, t_0)U(t_1, t_0)^\dagger = U(t_1, t_0)^\dagger U(t_1, t_0) = \mathbb{1}$ . By substituting this solution, and the dual of the Schrödinger equation which pertains to how the bras evolve in time into the definition of the density matrix, we find that the time evolution for a density matrix can be expressed in the form

$$(1.15) \quad i \frac{d}{dt} \rho = [\hat{H}, \rho]$$

where  $\hat{H}$  denotes the Hamiltonian of the system and the square brackets denote the commutator operation,  $[A, B] = AB - BA$ . We have now the Schrödinger equation for density matrices, also called the Heisenberg equations of motion for a quantum system.

Measurements in the quantum theory of mixed states can be well described with a rule, derived from postulate four of quantum theory, called the Born rule.

**Definition 1.4.** *The Born rule states that measurement in quantum mechanics can be achieved by considering a set of elements acting on the Hilbert space  $\mathcal{H}$  which are positive semi-definite matrices  $A_i$  that sum to the identity,  $\sum_i A_i = \mathbb{1}$ . Such a collection of operators are called a POVM. Also, by virtue of being positive semi-definite, the POVM elements may be further deconstructed to  $A_i = \sum_j K_{ij}^\dagger K_{ij}$ . The elements  $K_{ij}$  are often called Kraus operators. The Born rule itself is the rule*

$$(1.16) \quad p(i) = \text{Tr}(\rho A_i)$$

where  $p(i)$  is the probability to observe the measurement outcome  $i$ ,  $A_i$  is the POVM element associated to this measurement outcome and the state after measurement is given by the update rule

$$(1.17) \quad \rho \rightarrow \rho_i = \sum_j \frac{K_{ij}^\dagger \rho K_{ij}}{p(i)}.$$

We have seen how to connect states to density matrices. How can we go the other way? In order to connect density matrices to states, we can appeal to the purification principle. Let  $\rho$  be a density matrix acting on a Hilbert space  $\mathcal{H}_A$  (of finite dimension, the infinite dimensional case is well studied). To connect paradigms, we will need to define the partial trace operation. Then it is possible to construct a second Hilbert space and pure state  $|\phi\rangle\langle\phi| \in \mathcal{H}_A \otimes \mathcal{H}_B$  such that  $\rho$  is the partial trace of the pure state with respect to  $\mathcal{H}_B$ .

**Definition 1.5.** *The partial trace of an operator  $E$  over  $\mathcal{H}_A \otimes \mathcal{H}_B$  with respect to  $B$  is given by*

$$(1.18) \quad \text{Tr}_B[E] = \left( \sum_i \mathbb{1}_A \otimes \langle i|_B \right) E \left( \mathbb{1}_A \otimes |i\rangle_B \right)$$

where the set of vectors  $\{|i\rangle_B\}$  forms an orthonormal basis of the Hilbert space  $\mathcal{H}_B$  and  $\mathbb{1}_A$  is the identity operator on the system  $\mathcal{H}_A$ .

In order to see why this operation is important, consider a density matrix  $\rho$  defined on a Hilbert space  $\mathcal{H}_A \otimes \mathcal{H}_B$ . For an observable  $M$  on the system  $A$ , and the corresponding observable on  $A$  and  $B$  being given by  $M \otimes \mathbb{1}$ , it should be important that any assignment of a state onto  $A$  only gives identical measurement statistics. Indeed

$$(1.19) \quad \text{Tr}[(M \otimes \mathbb{1})\rho] = \text{Tr}[M\rho_A]$$

where we define  $\rho_A = \text{Tr}_B(\rho)$ . Using the tensor product and partial trace we can combine and ignore systems arbitrarily in quantum mechanics. We note that the Heisenberg equations of motion may appear to be violated if we do not observe the full Hilbert space in which a quantum system is defined, and instead only observe a fraction of the system. In order to account for this, we introduce the notion of a *quantum channel*. Quantum channels, much like their classical counterparts in Shannon theory, are useful for describing evolutions in communication-type scenarios, however their scope of application is broad and covers the paradigm of open quantum systems to name just a single use case.

**Definition 1.6.** *A quantum channel is a mapping from density matrices to density matrices that is both completely positive and trace preserving. Any quantum channel has an operator sum representation;*

$$(1.20) \quad \rho \rightarrow C[\rho] = \sum_i M_i^\dagger \rho M_i$$

where the  $M_i$  are operators satisfying  $\sum_i M_i^\dagger M_i = \mathbb{1}$ .

### 1.1.4 Thermodynamics

Thermodynamics is a theory that is often seen as apart, or distant from other physical theories such as general relativity, classical mechanics, or indeed quantum theory. Perhaps the reason for this is thermodynamics makes no attempt to understand mysteries in the natural world, only how to best exploit the natural world for our own gains [12, 13, 14, 15]. Thermodynamics makes no attempt to unearth microscopic details of physical systems, or make statements about how physical systems will evolve given a certain initial condition. Note here I differentiate between thermodynamics as a concept, and certain topics that often fall within it, such as the heat equation, which certainly do make assertions about how systems will evolve in time. Thermodynamics is instead a collection of statements about permissible transformations between systems, and which transformations among these are easy to implement and which are hard.

Thermodynamics can be summarised by the following four laws, in which the character of thermodynamics as a physical theory is found. Note that all of these laws, are more in the spirit of empirical facts that are observed within nature, and can be proven to hold as laws within regimes such as ideal gasses in a heat engine. The ubiquity of application of these facts are the reason they are called laws;

**Definition 1.7.** *The zeroth law of thermodynamics can be stated as if two systems are both in thermal equilibrium with a third system, then they are in thermal equilibrium with each other. Note that as with all the laws, there are multiple ways to state them, and all imply the same physical facts on a system. This law implies a transitive relationship among the temperatures of multiple physical bodies.*

**Definition 1.8.** *The first law of thermodynamics is a statement of the conservation of energy. In a closed system, the change in internal energy of the system ( $\Delta U$ ) is equal to the difference between the heat flow ( $Q$ ) and the work done by the system on its surroundings ( $W$ ). Equivalently,  $\Delta U + Q + W = 0$ . Note here, that we fix the sign convention of positive work done by the system on the surroundings.*

**Definition 1.9.** *The second law of thermodynamics tells us that the entropy of a system is necessarily increasing over time. We will use the Kelvin statement of the second law. It is impossible for a self-acting machine, unaided by any external agency, to convey heat from one body to another at a higher temperature.*

**Definition 1.10.** *For the third and final law of thermodynamics, we appeal to Nernst: It is impossible for any process, no matter how idealised, to reduce the entropy of a system to its absolute-zero value in a finite number of operations.*

From these four laws, it is possible to derive a great many results concerning thermodynamic systems. They forbid the construction of perpetual motion devices. They establish the Carnot theorem, which bounds the maximum efficiency achievable by any heat engine that operates between a hot and cold reservoir in terms of the temperatures of said reservoirs. They establish well the physical properties of materials, such as the specific heat capacity or the coefficient of thermal expansion in terms of simple system parameters such as temperature or volume and the response of the system under changing external parameters.

In an attempt to align quantum theory, and specifically quantum information theory with thermodynamics, Von-Neumann designed the following thought experiment to determine the entropy of a density operator [12]. Consider a box of ideal gasses initially in some container of volume  $V$ . Initially, the density operator is in an ensemble over  $n$  states  $\rho^{\otimes n}$  with  $\rho = \sum_k p_k |\phi_k\rangle \langle \phi_k|$ . The thought experiment consists of attempting to transform the state to a pure state  $|\phi_1\rangle^{\otimes n}$ , by means of a reversible process. The transformation consists of three steps.

1. First, the different species of gas are separated into states  $|\phi_1\rangle, \dots, |\phi_n\rangle$  inside different boxes, all with a volume  $V$  identical to the original container. This transformation could be achieved with, for example, semi-permeable membranes. No work is done and no heat is exchanged.
2. Each state undergoes an isothermal compression into a volume  $V_k = p_k V$ . The mechanical work done in this process is  $W_k = n p_k \ln \frac{V_k}{V} = p_k \ln p_k$ , in units where  $k_B T = 1$ . Since the

compression is isothermal the heat flow is the negative of the work, and hence the entropy increase per particle is  $\Delta S = \sum_k p_k \ln p_k$ .

3. Each box is now set to  $|\phi_1\rangle$  by means of a unitary transformation, which is possible since the states have been separated into classically distinguishable entities.

In this three step process, the entropy of the final state, being pure (and so carrying no uncertainty about which quantum state we have) is zero. Hence, the entropy of the initial system, is equal to the change in entropy and so  $S(\rho) = -\text{Tr}(\rho \ln \rho)$ . This quantity is known as the Von-Neumann entropy of the state.

**Definition 1.11.** *The Von-Neumann entropy of a quantum state represented by a density matrix is given by*

$$(1.21) \quad S(\rho) = -\text{Tr}[\rho \ln \rho].$$

Immediately, this places thermodynamics and quantum theory in contention with one another. Thermodynamics posits that on one hand, the entropy of the universe is increasing monotonically, yet on the other, quantum theory says that the entropy of the universe undergoing unitary evolution does not change. This can be seen algebraically

$$(1.22) \quad S(U^\dagger \rho U) = \text{Tr}(U^\dagger \rho U \ln U^\dagger \rho U) = \text{Tr}(U^\dagger \rho U U^\dagger \ln(\rho) U) = \text{Tr} \rho \ln \rho = S(\rho)$$

where we have utilised the cyclic symmetry of the trace and the identity that holds for unitary  $U$  and Hermitian  $\rho$ ,

$$(1.23) \quad e^{U^\dagger \rho U} = \sum_n \frac{(U^\dagger \rho U)^n}{n!} = U^\dagger \left( \sum_n \frac{(\rho)^n}{n!} \right) U$$

which implies that  $\ln U^\dagger \rho U = U^\dagger \ln(\rho) U$ . The resolution to this apparent paradox is, as we will see in chapter two, while the overall entropy may be unchanging under unitary evolution, the parts of the system we can actually observe (and we do only observe a finite subsection of the universe) may be increasing in entropy as correlations build between unseen, or unobserved degrees of freedom. The Von-Neumann derivation itself can also be seen as incomplete - it has violated the third law of thermodynamics. The third law itself is very interesting in the context of quantum theory. A recent article [16] has argued that ideal projective measurements should be impossible, precisely because they, in a single step, permit one to generate a pure, zero entropy state. One could write an entire thesis on the contentious issue of quantum thermodynamics, and it would be a vast and brilliant study. Here, we restrict ourselves to mention these issues only in passing so that we may focus on other, more tangible aspects of quantum thermodynamics. In chapter two, my main chapter on thermodynamics, we visit the issue of work extraction in the quantum world.



There is also another aspect in this contention which is interesting. At some level, thermodynamics has been viewed as more fundamental than all other physical theories. Take the famous quote from Eddington [17] - "If your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation". Yet, once we specify the equations of motion for a system (such as the Schrödinger equation), we lose the privilege of utilising a thermodynamic description. The equations of motion are viewed as fundamental and exact, and thermodynamics is inherently approximate. This is inherent to the tenet of 'initial conditions plus equations of motion' viewpoint of modern physics. David Deutsch cites this as one of his motivations for his developments of constructor theory [18], and we will see in the later chapters of this work how another different attempt to get around the initial conditions plus equations of motion mantra is compatible with thermodynamics.

All of the above discussion on entropies takes place in the identically and independently distributed (i.i.d.) quantum systems regime, also called the asymptotic regime. One imagines they have many copies of a quantum system and are interested in what they can do with it, with the entropy being related to some average (perhaps an average energy) in these tasks. We also note [19] that surprisingly, the averaging also holds when one considers a single quantum system. Quantum information theory has its success rooted in its applicability and many of the tasks associated with this regime are not what one would typically call thermodynamic. Examples include the Schumacher compression [20], quantum hypothesis testing [21] and entanglement distillation [22]. Despite their operational interpretation being a far cry from temperature and steam engines, figures of merit in these types of tasks are entropic in nature (possibly depending on extensions to entropy between systems such as the mutual information).

There is another regime, where one imagines they hold only a single quantum system and are interested in deterministic (or almost deterministic) transformations thereof. Figures of merit in these types of tasks are essentially all the quantum Renyi divergences [23, 24, 25, 26].

### **1.1.5 Bell Inequalities**

In this section we will clarify what is so special about quantum entanglement, we will explain what is meant by a Bell inequality and we will encounter our first physical principle that is not a conservation law - the principle of No-signalling.

In the early 20th century, physics seemed to be at war with itself. Einstein, Podolski and Rosen [27] cooked up a paradoxical situation by making use of the entanglement in quantum theory whereby an observer could in principle know both the position and momentum of a particle, by making measurements of a second particle that is entangled with the first. This is in clear violation with the Heisenberg uncertainty principle. Indeed Bohr [28] wrote long letters to Einstein arguing for the Heisenberg viewpoint - that the thought experiment was misleading. It was not until the groundbreaking 1964 paper by John Bell [29], who presented results concerning measurement of the spin of an entangled quantum system that clearly demonstrated there

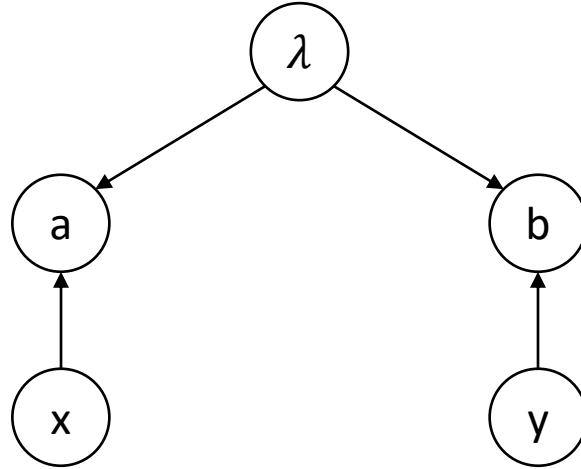


Figure 1.2: The hidden variable common cause explanation for the Bell scenario. There exist experiments in quantum theory that cannot be explained in this manner.

existed situations in quantum physics that were inconsistent with the EPR reasoning.

In a Bell scenario, pairs of systems are prepared at some location in space-time. They are then separated by a great distance sufficient to impose space-like separation on the subsequent measurements of the system pairs. The natural explanation for correlations among the measurement outcomes is that they are both a function of the measurement settings chosen at each side of the great distance, as well as of a common cause located in the joint past of the pair of systems. Bell's famous result is one which rules out this possibility. A hidden classical variable cannot explain the correlations observed when one conducts this experiment on pairs of quantum systems. A Bell inequality is exactly a hyper-plane that separates correlations that are achievable in models with a classical common cause and those that are not. Many experiments have been devoted in the time since 1964 to observing such correlations [30]. And, up-to loopholes [31, 32, 33] which will not enter the discussion of this thesis, they have observed correlated measurement outcomes which lie on the 'classically forbidden' side of a Bell hyper-plane.

The most famous and prototypical scenario where one can construct a Bell inequality is in the Clauser, Horne, Shimony, and Holt scenario [34]. We take the space like separated parties to be measuring a two-valued outcome that are assigned  $\pm 1$ , with detector settings chosen from  $x \in X$  for the first party and  $y \in Y$  for the second party. Let  $E(x, y)$  be the product of the outcomes observed under the condition the detector settings are  $x$  and  $y$ . Any classical model that relies on a common cause between the variable outcomes can be written [35],

$$(1.24) \quad E(x, y) = \int A(a, \lambda)B(b, \lambda)p(\lambda)d\lambda$$

where  $A$ , and  $B$  are the average values of the outcomes of measurements on either side of the set-up, conditioned on the settings of the measurement device being chosen as  $x$  and  $y$ , and  $\lambda$  is

the hidden variable, distributed according the classical probability  $p(\lambda)$ . By taking differences

$$\begin{aligned}
 E(x, y) - E(x, y') &= \int [A(x, \lambda)B(y, \lambda) - A(x, \lambda)B(y', \lambda)]p(\lambda)d\lambda \\
 &= \int [A(x, \lambda)B(y, \lambda) - A(x, \lambda)B(y', \lambda) \pm A(x, \lambda)B(y, \lambda)A(x', \lambda)B(y', \lambda) \mp A(x, \lambda)B(y, \lambda)A(x', \lambda)B(y', \lambda)]p(\lambda)d\lambda \\
 (1.25) \qquad &= \int A(x, \lambda)B(y, \lambda)[1 \pm A(x', \lambda)B(y', \lambda)]p(\lambda)d\lambda - \int A(x, \lambda)B(y', \lambda)[1 \pm A(x', \lambda)B(y, \lambda)]p(\lambda)d\lambda
 \end{aligned}$$

By taking absolute values and application of the triangle inequality

$$\begin{aligned}
 |E(x, y) - E(x, y')| &\leq \\
 (1.26) \qquad &\leq \left| \int A(x, \lambda)B(y, \lambda)[1 \pm A(x', \lambda)B(y', \lambda)]p(\lambda)d\lambda \right| + \left| \int A(x, \lambda)B(y', \lambda)[1 \pm A(x', \lambda)B(y, \lambda)]p(\lambda)d\lambda \right|
 \end{aligned}$$

the integrands  $A(x, \lambda)B(y, \lambda)[1 \pm A(x', \lambda)B(y', \lambda)]p(\lambda)$  and  $A(x, \lambda)B(y', \lambda)[1 \pm A(x', \lambda)B(y, \lambda)]p(\lambda)$  are non-negative. Using this fact we write the right hand side of this inequality as:

$$(1.27) \qquad \int |A(x, \lambda)B(y, \lambda)||[1 \pm A(x', \lambda)B(y', \lambda)]p(\lambda)d\lambda| + \int |A(x, \lambda)B(y', \lambda)||[1 \pm A(x', \lambda)B(y, \lambda)]p(\lambda)d\lambda|$$

as they are the products of numbers which are either 1, 0 or  $-1$ ,  $|A| \leq 1$  and  $|B| \leq 1$ , then,

$$(1.28) \qquad \leq \int [1 \pm A(x', \lambda)B(y', \lambda)]p(\lambda)d\lambda + \int [1 \pm A(x', \lambda)B(y, \lambda)]p(\lambda)d\lambda$$

the integral  $\int p(\lambda)d\lambda = 1$  by definition that  $p$  is a continuously valued probability. Therefore this quantity is nothing more than

$$(1.29) \qquad 2 \pm \int [A(x', \lambda)B(y', \lambda) + A(x', \lambda)B(y, \lambda)]p(\lambda)d\lambda = 2 \pm (E(x', y') - E(x', y))$$

and putting the left hand side of the equation back into the formula and after some algebraic manipulation we find that

$$(1.30) \qquad 2 \geq |E(x, y) - E(x, y') + E(x', y') + E(x', y)|$$

which is the CHSH inequality (*ibid*). The Bell result is to conduct an experiment which violates this inequality by making use of measurements of particles in a singlet state [36], to achieve a value for  $|E(x, y) - E(x', y) + E(x', y') - E(x, y')|$  of up-to  $2\sqrt{2}$ , a figure which is known as Tsirlson's (or Cirel'son's) bound [37]. Therefore, we must conclude that there is no model of classical common cause with hidden variables which can account for the correlations between results of measurements in this quantum experiment.

With this in mind, can we construct a physical principle that quantum particles must obey in distributed measurement scenarios? Clearly, many experiments have shown that our candidate

cannot be ‘nature does not violate Bell inequalities’ because we can show that it does. Consider once again the problem of two parties making measurements of a system. The parties are space-like separated during the experiment and they make their measurements such that there is not time enough for a photon to fly between them. Label the parties A, for Alice and B, for Bob. Let the types of measurement that Alice will make be indexed by a random variable  $x$ , and let the types of measurements Bob will make be indexed by  $y$ . These variables are generated by each party at the start of the experiment and not shared between parties. The outcomes of these measurements are labelled  $a$  for Alice and  $b$  for Bob. The distribution over the outcomes is represented by the conditional probability distribution  $p(a, b|x, y)$ . By the rules of probability theory, we can marginalise over Bob’s outcomes so that  $p_A(a|x, y) = \sum_b p(a, b|x, y)$ , where  $p_A$  is the conditional probability observed by Alice. Let us assume they run this experiment many times  $N$  and the distance between them is so great that a photon flying between the parties at the instance of the first experimental run will not have finished its journey by the time the  $N$ ’th experiment is concluded.

Let us pause for a moment and think about the possible distributions we can achieve in such an experiment. More than that, let us appeal to the physical set-up to do so. By picking  $y = y'$  for the first  $N/2$  runs of the experiment Alice is able to accurately and obeying good statistical practice construct the distribution  $p_A(a|x, y')$  for all values of  $x$  she may pick. After the  $N/2$ th run, Bob decides to switch his input to  $y = y''$ . Alice is then able, using the remaining experimental runs construct the distribution  $p_A(a|x, y'')$ . If it is true that  $p_A(a|x, y') \neq p_A(a|x, y'')$  for some values of  $x$ , Alice is able to detect this difference and conclude that Bob has changed his input. This means that Bob, despite the great distance between them is able to instantaneously signal to Alice! Clearly this is incompatible with the special theory of relativity. Instantaneous action at

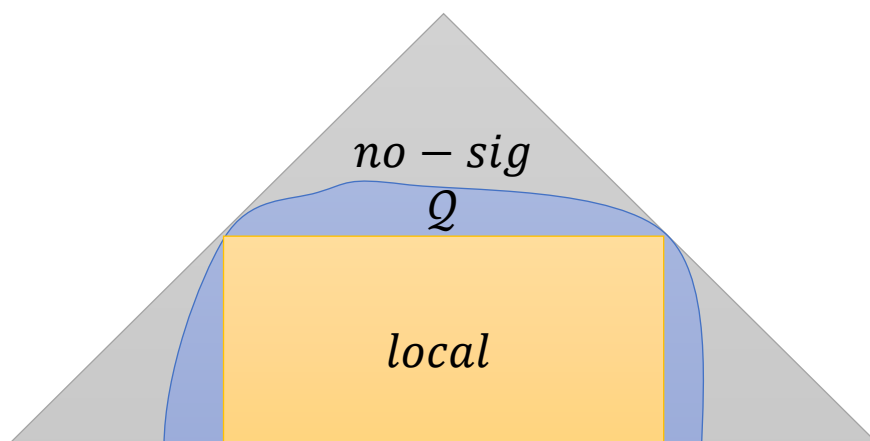


Figure 1.3: A 2D slice through the Local and No-signalling polytopes. The set of quantum achievable correlations  $\mathcal{Q}$  sits between these two polytopes. PR boxes sit at the vertices of the No-signalling set. Local is the set of all correlations that can be written with a classical common cause model.

a distance is prevented in this theory, and so if we want to believe, as in relativity, that there is a local structure to the universe in which we live we demand that  $p_A(a|x, y') = p_A(a|x, y'')$ . By symmetry we impose the physical principle on Bob's side too, and so arrive at our physical principle - the No-signalling conditions

**Definition 1.12.** *The No-Signalling conditions for bipartite probability distributions are that*

$$(1.31) \quad \sum_b p(a, b|x, y) = p_A(a|x, y)$$

$$(1.32) \quad \sum_b p(a, b|x, y) = p_B(b|x, y)$$

*are independent of  $y$ , and  $x$  respectively.*

No-signalling conditions are obeyed by all probability distributions that are achievable in classical theory or quantum theory. There exist No-signalling compatible probability distributions that are neither achievable by quantum theory or classical theory. The most famous example of these are the Popescu Rohrlich Correlations [38]- which are highly non-local but none the less are compatible with the No-signalling principle. The prototypical example of Popescu Rohrlich correlations are those of the PR box, where

$$(1.33) \quad p(a, b|x, y) = \frac{1}{2} \delta_{xy, a \oplus b}$$

which can be shown to take a CHSH value of 4 by substitution into (1.30). The gaps between the correlations that obey NS and those achievable in quantum theory has lead towards research into more exotic physical principles than the No-signalling conditions in recent years, most notably information causality [39].

### 1.1.6 The Problem of Defining Time

In quantum theory, position and momentum are Hermitian operators and form a conjugate pair through the Fourier transform. The corresponding commutation relation that matches them is given by  $[\hat{x}, \hat{p}] = i$ . We might similarly expect that there exist canonical commutation relations between energy and time in quantum theory. Momentum is the generator for shifts in position, and vice versa. So given that energy acts as the generator of time translations through the Schrödinger equation, where is the commutation relation for energy and time? An objection due to Pauli [40] shows this not to be possible, because no time operator exists. The argument goes as follows. Assume one such operator exists,  $[\hat{T}, \hat{H}] = i$ . As  $\hat{T}$  is Hermitian the operator  $\hat{U}_\omega = e^{-i\omega\hat{T}}$  is unitary. A formal power expansion of this operator commutator yields

$$(1.34) \quad [\hat{U}_\omega, \hat{H}] = \sum_k \frac{(-i\omega)^k}{k!} [\hat{T}^k, \hat{H}] = -\omega \hat{U}_\omega.$$

As direct consequence of this we have that the spectrum of the Hamiltonian is the entire real line. To see this, consider an energy eigenstate  $|e\rangle$ . Then we find that

$$(1.35) \quad \hat{H}\hat{U}_\omega|e\rangle = (e + \omega)\hat{U}_\omega|e\rangle$$

and so that  $\hat{U}_\omega|e\rangle$  is an eigenstate of  $\hat{H}$  with eigenvalue  $(e + \omega)$ . At no point do we assume that  $\omega$  need be positive, the statement about the unboundedness of the energy spectrum of the Hamiltonian follows. This is a radical problem for contemporary physics. We could repeat the above proof with the position and momentum operators and arrive at no such dire consequence. In quantum theory there is really no restriction on the negativity of position or momentum. One might like to think that objects can be as far left as they can be far to the right. However, for energy, we impose the restriction that energy must be a positive quantity. At the very least, we would like energy to be bounded from below, which the above proof forbids.

A variety of objections to the Pauli argument have cropped up in the literature. For example, it has been shown that an external system which can act as a clock gets around the Pauli argument [41]. However, the time operator in this formalism is not conjugate to the system Hamiltonian and as such is not the generator for energy shifts in the system degrees of freedom. Despite this, it is an important first step should one wish to revoke Pauli's objections. This Page-Wooters approach to time in quantum theory [42] is to consider that time is emergent from an underlying entanglement between systems in the universe. While grandiose, this idea cropped up in a different context, the quantum theory of gravity, and did form at least part of the basis for the formation of the Wheeler De-Witt approach to quantum gravity [43], which in turn inspired loop quantum gravity by Rovelli [44]. Loop quantum gravity is one of the largest and best attempts to arrive at a full theory of quantum gravity, the other being string theory, with development of such theories ongoing. The Page-Wooters approach has also recently been revisited, and shown to be both resistant to criticisms of ambiguity, and updated in a modern fashion [45].

We also note that while it may not be possible to define a universal time operator in quantum mechanics, time of arrival and time of flight operators [46] can be defined and shown to possess all of the desired properties they would need to carry within their remit of application. I want to stress that it is not forbidden to construct any valid operator that has time in it, but a time operator conjugate to the system Hamiltonian is a no-go. An example, quantum theory has been shown to apply when calculating the theory of high precision clocks [47, 48, 49], such as clocks based on the processions of a caesium atom [50]. Interestingly, how good a quantum clock can be at telling the time comes, much like the Heisenberg uncertainty, with a bound that typically grows the longer the clock is left to run. More generally, we can appeal to the Wigner-Araki-Yanase theorem [51], which states that there is a fundamental limit to how well operators that do not commute with conserved quantities can be measured. None-the-less, the absence of a time operator in quantum theory does pose significant philosophical challenges for anyone who wishes to interpret quantum theory.

It is not just a problem to define time in quantum theory, which can be explained away by the fact that time enters as a parameter in the theory. A closely related concept, causal order also poses problems. The following demonstrative example I attribute to Oppenheim [52]. Consider the case of two particles travelling on a line. Both are initially localised to the right of the origin and are travelling to the left. We wish to ask which of the particles arrives at the origin first. The Hamiltonian for the system and the measurement apparatus is given by

$$(1.36) \quad \hat{H} = \frac{\hat{p}_x^2}{2m_x} + \frac{\hat{p}_y^2}{2m_y} + \hat{H}_{int}$$

where  $\hat{p}_x$  and  $\hat{p}_y$  are the momentum operators of the particles and  $\hat{H}_{int}$  is some interaction Hamiltonian that achieves the measurement. A good interaction Hamiltonian is given by

$$(1.37) \quad \hat{H}_{int} = \alpha \delta(x) \theta(-y)$$

where  $\delta(x)$  is understood to mean the Dirac delta distribution on  $x$  and  $\theta(y)$  is the Heaviside step function on  $y$ . The parameter  $\alpha$  tends to infinity. Classically, this is a good measure for the order in which the particles arrive by the following logic. If the  $y$  particle arrives before the  $x$  particle, the potential for  $x$  at the origin is switched on and the  $x$  particle is reflected back. After the scattering, the  $x$  particle will be found to the right and the  $y$  will be found to the right. If instead the  $x$  particle arrives first, then the switching on that comes when  $y$  passes the origin does not effect the trajectory of the  $x$  particle (the derivative of this potential is 0 in the region that the  $x$  particle is localised in), the ‘door closes’ only when the particle has already traversed the threshold and so at the end of the experiment both particles will be found to the left of the origin. Quantum mechanically however, this logic fails. To see why, we decompose the space into a position representation where we plot the wavefunction over the four quadrants of a two dimensional plane, with each coordinate representing the position of one of the particles.

In the classical case, after the experiment, following the logic above the particle will be found with certainty in quadrants 3 or 4. Quantum mechanically, there exist solutions in which the particle can be found in quadrants 1 and 2 as well. We see that the solution to the problem is equivalent to the solution for particles scattering off a thin edge. Handy, since this problem has a known solution. Assume that the boundary condition is such that (switching to polar coordinates)  $|\psi(r, \phi)\rangle = 0$  for  $\phi = 3/2\pi$ . Modelling the two particles within the plane wave approximation allows us to write the solution for an incident plane wave which makes an angle  $\phi_0$  with the  $x$  axis,

$$(1.38) \quad \psi(r, \phi) = \frac{1}{\sqrt{i\pi}} (e^{-ikr \cos \phi - \phi_0} \Phi(\sqrt{2kr} \cos \frac{\phi - \phi_0}{2}) - e^{-ikr \cos \phi + \phi_0} \Phi(-\sqrt{2kr} \sin \frac{\phi + \phi_0}{2}))$$

where  $\Phi$  is the Gaussian error function. Asymptotically,

$$(1.39) \quad \psi(\phi) = \begin{cases} e^{-ikr \cos \phi - \phi_0} + f(\phi) \frac{e^{ikr}}{\sqrt{r}} & -\phi_0 < \phi < \phi + \pi \\ e^{-ikr \cos \phi - \phi_0} - e^{-ikr \cos \phi + \phi_0} + f(\phi) \frac{e^{ikr}}{\sqrt{r}} & -\phi_0 > \phi > \frac{\pi}{2} \\ f(\phi) \frac{e^{ikr}}{\sqrt{r}} & \pi - \phi_0 < \phi < \frac{3\pi}{2} \end{cases}$$

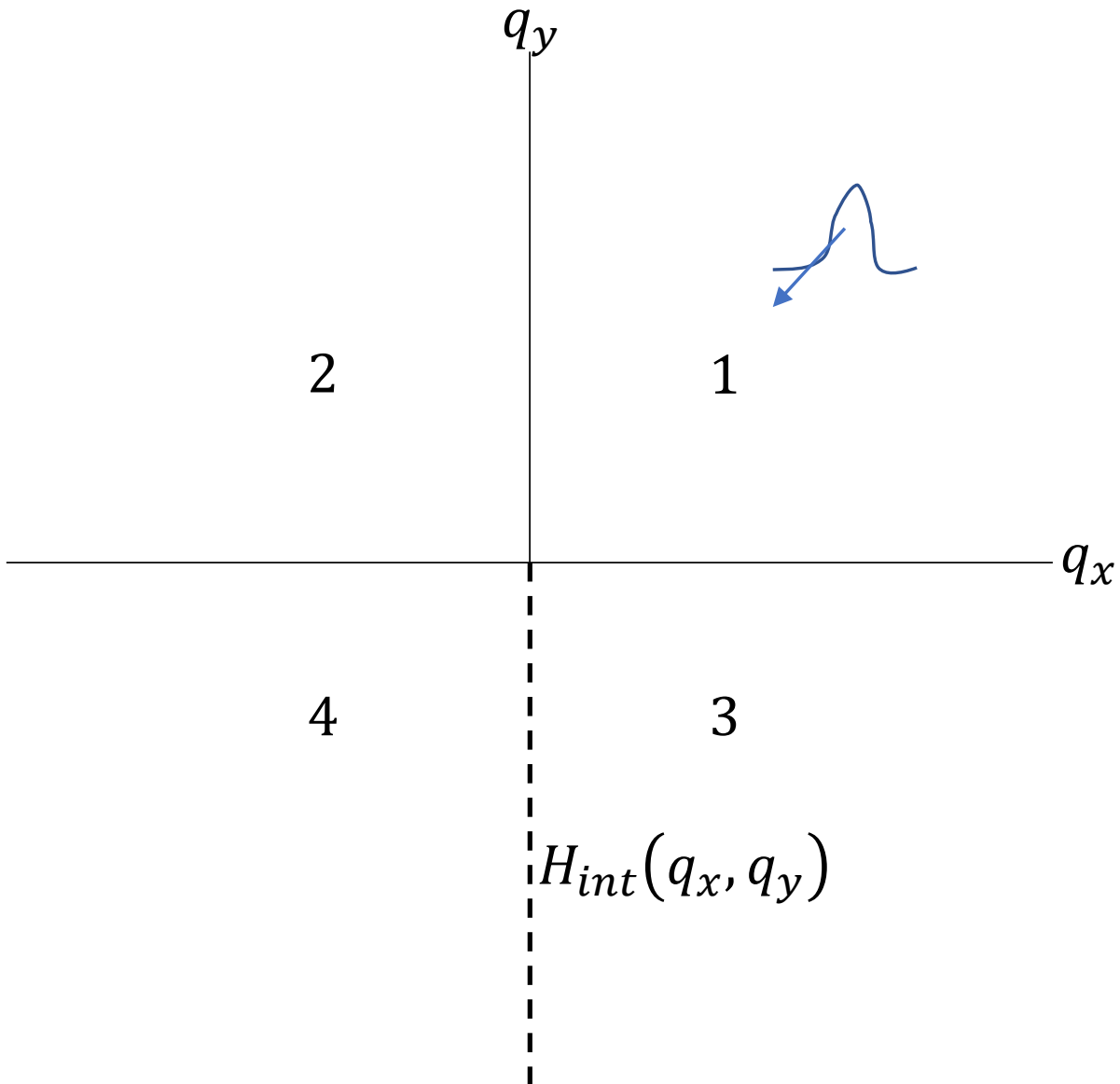


Figure 1.4: Before the scattering, the particles are in quadrant one, and travelling towards the negative  $x$  and the negative  $y$  axis. Afterwards, classically, the particles will be in either quadrant three or four. Quantum mechanically there are solutions where the particles are found back in quadrant one, or even in quadrant two!

with

$$(1.40) \quad f(\phi) = -\sqrt{\frac{i}{8\pi k}} \left( \frac{1}{\sin 1/2(\phi + \phi_0)} + \frac{1}{\cos 1/2(\phi - \phi_0)} \right).$$

The above approximation becomes invalid when  $\cos(\frac{\phi - \phi_0}{2})$  or  $\sin(\frac{\phi + \phi_0}{2})$  is close to zero. None the less, we can by inspection observe that the particles can be found in the classically forbidden regions 1 and 2. By defining a failure cross section (failure taken to mean we are unable to



distinguish between the order of arrival of the particles) we find

$$(1.41) \quad \sigma_f = \int_0^{2*\pi} |f(\phi)|^2 d\phi = \frac{1}{k} \frac{1}{\cos \phi_0/2}$$

which is the wavelength of the incident particle multiplied by an angular dependence. At best we can say if the particles are coincident on the scattering potential to within

$$(1.42) \quad \delta x > 2/k,$$

the measurement has a good chance of failing. By relating plane waves back to two particles, we can use the approximate relation

$$(1.43) \quad \delta t \approx \frac{m\delta x}{k}$$

to make a connection between the spatial and time co-ordinates of the particles. Combining these relations gives us a form of the ‘Heisenberg uncertainty relation for energy and time’

$$(1.44) \quad \delta t > \frac{1}{E}.$$

which is to say, we cannot use this procedure to reliably determine the causal order of arrival if the particles arrive a time shorter than  $1/E$  of each other, with  $E$  interpreted as the total kinetic energy of the particles. Clearly, the plane wave approximation is invalid since we claimed that the particles are both initially to the right of the origin, an assertion one cannot make with plane waves. To present analytical solutions and not resort to numeric calculations and expectation values instead of actual energies it is one we make here for brevity. One can repeat the analysis for particles with a large spread of momentum eigenstates and find that the measurement will almost always fail for incident particles with  $\delta t < 1/\bar{E}$  where  $\bar{E}$  is the minimum typical energy.

Although this example is very nice, the fact that energy and time do not form a ‘proper’ uncertainty relation has been explored at length [53, 54, 55]. Important distinctions come in the cases where the Hamiltonian for the system is known versus unknown, in the known case a measurement of energy can be performed in an arbitrarily short time, however if the Hamiltonian is not known then to obtain an uncertainty in energy of  $\Delta E$  the system measurement will take at least  $\Delta t$ , where  $\Delta t \Delta E \geq 1$ . There are no such nuances in the ‘proper’ Heisenberg uncertainty relations for position and momentum for example, these are facts of quantum mechanics.

The reason for including this example in this thesis was to make apparent that the issues in quantum mechanics with time go much deeper than can be explained away with the mantra of ‘time being a parameter’ in quantum theory. Certainly, time in quantum theory is a strange object, but causal order of events more so. The possibility for the causal order of events to be placed into a superposition and as such exhibit quantum mechanical interference will be a main topic in this thesis.

## 1.2 Causality

We now move on to the second section of the introduction to this thesis, which concerns the topic of causality, and is where we will meet our second physical principle. Causality is one of the widely studied topics, and has a long history dating back to antiquity, where Aristotle claims ‘we think to have knowledge of a thing only when we have grasped its cause’. What though, does it mean for one event to be the cause of another. One is tempted to adopt the view that one event may be classed the cause of another when ‘tinkering with the former event can change the latter’. Before even moving to the quantum mechanical case this poses problems, although the idea of the physical principle we want to motivate is embedded in this statement. Problems of course occur because it is not possible to observe the change in a single event. Indeed, if the farmer sells the rooster in the afternoon of the first Friday of June, he will not hear the ‘cockadoodle-doo!’ on Saturday morning, and he may be satisfied that his selling of the bird is what has caused him to have a restful sleep. However, a Saturday in June does not exist where the rooster is still on his farm, and so the farmer is spuriously associating the experiences of waking up early in the morning of every other Saturday morning throughout the year past with the present. Perhaps, should the farmer have not sold the rooster, a crafty fox may have taken the opportunity to make use of the solstice night of Friday evening to hunt the rooster - and the farmer, none-the-wiser to the murderous event would have slept in until the late morning on Saturday anyway - Its an edge case for sure, but can we now say the farmer selling the rooster *caused* his restful sleep? What the farmer is doing in his head is making a statistical comparison between the past and the future, and more over, making causal inference about the underlying model which causes or does not cause him to leave his bed. The formalism of such models, the process by which the farmer can make inference and the physical principle underpinning these events are the subject of this section.

A warning, this section is devoted to learning a more complete picture of the modern causality literature, and I have endeavoured to make the rest of this thesis read smoothly should one wish to ignore this section. When we begin to prove claims, we will appeal to our own notation and ideas. However, I feel that the picture one wishes to paint is much more complete if the literature on classical causality is at least covered.

### 1.2.1 History

We begin our discussion of causality with a short history from antiquity to state-of-the-art methods. One of the earliest statements to which one can draw comparisons between what we would today call causal inference and what philosophers of the past would, is due to Democritus (who, incidentally also made statements about the granularity of nature and perhaps the work in this thesis can ultimately be attributed to!) who proclaimed that ‘I would rather discover one causal law than be King of Persia’. We don’t know for sure Democritus’ personal views on

becoming the holder of one of the most prominent titles in all of the ancient world, but I read the statement thus - 'it is *hard* to discover the true causes of events'. Many across the ancient world were betrayed, bullied, deceived, and ultimately murdered in the hope of merely sitting aside the great king, Democritus tells us despite all this palace intrigue, it is none the less a harder task to determine if the rooster not raising the farmer was caused by the farmers own action.

The next major philosopher who discusses causation in a meaningful way is Hume [56]. Hume defines two conditions for an event to be the cause of another, and he defines the actual cause to be a relation between objects that we employ in our reasoning in order to yield less than demonstrative knowledge of the world beyond our immediate impressions. The two conditions are given as

1. An object precedent and contiguous to another, and where all the objects resembling the former are placed in like relations of precedency and contiguity to those objects that resemble the latter.
2. An object precedent and contiguous to another, and so united with it, that the idea of the one determined the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other.

Obscured by Humian notions of the mind, Hume none the less here postulates a valuable insight. He simply says that causes precede effects, according presumably to some notion of timing, and that there is a relation between them where knowledge of the former might enable us to imagine the latter better. There is also here some notion of 'all other things being equal' which eludes to the farmers fox, who has not visited the farm this year and so the farmer is implicitly assuming he will not do so on the first Friday night in June either. Causality was discussed by a variety of philosophers from Hume, however none up until the 20th century had more meaningful insights than him.

It is at the next point in history we introduce probabilities into nature. R.A Fisher was a statistician who formalised the modern gold standard for research - a 'controlled experiment'. Fisher's contribution was to observe a simple fact about regression towards the mean. Suppose you are running a trial for the effectiveness of a drug on the prevalence of a heart condition. You want to test the drug on those which are at risk for some heart problem, and you measure as a proxy, say, their cholesterol level. The drug is meant to reduce the level of cholesterol in the patients blood.

Now, if you run a trial where you survey the cholesterol level in the blood of a sample of the the general population. You then pick those patients with the highest cholesterol level to test the drug on. You administer the drug, and then six weeks later the patients have their blood cholesterol level recorded again. It is a fact, that given a large sample size, that upon the second recording, the average blood cholesterol level in the population you survey will have gone down. Hurrah, you proclaim, and deem the drug effective in combating high cholesterol.

The problem is, that the average would have fallen had you not administered the drug anyway, because you picked a biased sample to test. This phenomena is called regression towards the mean, and has been observed and the cause of many misinterpretation of facts through the years - in the case of how 'negative reinforcement' works for fighter pilots but no-one else [57], or the simple fact that two abnormally tall parents will not produce a child with an even more extreme height in general.

Fisher's ingenious idea was to introduce a control group, which are picked among the same sample as the test group (one way to make this true is to simply randomise who will be administered the treatment in the picked sample group) so that you could compare the effectiveness of the drug against the statistical effect of regression towards the mean. Under the circumstances, if you observe the drug group cholesterol reduce more than the control group, you would say that the drug has caused the reduction of cholesterol. Nuances in this are numerous.

Randomised controlled experiments are still viewed as the gold standard today in many fields, however alternatives to randomised controlled experiments are being proposed, first in modern times by Pearl [58], but by others.

### 1.2.2 Formalism

In this section we will very briefly describe some modern methods for inference, by making heavy use of directed acyclic graphical structures. Before this, we will need to define what it means for two probability distributions to be conditionally independent.

**Definition 1.13.** *Let  $V = V_1, V_2, \dots$  be a finite set of variables. Let  $p(\cdot)$  be a joint probability over these variables. Lets  $X, Y, Z$  be any three subsets of variables in  $V$ . The sets  $X$  and  $Y$  are conditionally independent given  $Z$  if and only if*

$$(1.45) \quad p(X, Y | Z) = P(X | Z)P(Y | Z)$$

which we will denote with the notation  $X \perp\!\!\!\perp Y | Z$ .

A graph  $G$  consists of nodes or vertices  $V$  and a set of edges  $E$ . let the vertices correspond to variables that appear in a probability assignment  $p$ . Edges will be assigned where there is a certain relationship between variables, to be clarified. Vertices connected by an edge are termed adjacent, and a graph will have a unique description in terms of these edge assignments, often stored in an adjacency matrix.

Each edge of a graph can be either undirected (no arrows), directed (a single arrow) or bidirected (two arrows pointing in both directions along the edge. If all edges in a given graph are directed we say that we have a directed graph. The skeleton of a graph is the graph constructed by removal of all directional structures along the edges. A path in the graph is a sequence of edges where the start of one edge is the final vertex of the preceding edge. A directed path is one such path that respects the edge orientation in the graph. A cycle is any path that terminates and

originates at the same vertices, and a directed acyclic graph (DAG) is any graph which contains so such directed cycles.

A parent (A) of a node (B) in a directed acyclic graph are those nodes which have directed paths from themselves (A) to the node (B). A child is a node (C) which has directed path to it originating from (A). The acyclic requirement prevents nodes from simultaneously being a child and a parent to another node.

Suppose that we have a probability  $p$  defined on  $N$  variables, each one corresponding to a node in the graph. We may order the variables arbitrarily by labelling them as  $X_1, \dots, X_n$ . By the chain rule of probability calculus we can decompose the joint distribution over all variables into conditionals for each variable,

$$(1.46) \quad p(x_1, \dots, x_n) = \sum_i p(x_i | x_1, \dots, \cdot_i, \dots, x_n)$$

where the  $\cdot_i$  indicates  $x_i$  is not present on the right hand side of the conditional.

We will next need the idea of Markovian parents in a distribution.

**Definition 1.14.** Let  $V = X_1, X_2, \dots$  be a finite set of variables. Let  $p(\cdot)$  be a joint probability over these variables. A set of variables  $PA_j$  are said to be Markovian parents to  $X_j$  if  $PA_j$  is the minimal set of predecessors that renders  $X_j$  conditionally independent of all its other predecessors;

$$(1.47) \quad p(x_j | pa_j) = p(x_j | x_1, \dots, x_{j-1})$$

This definition now allows us to uniquely specify a directed acyclic graph compatible with this probability distribution to it. We draw arrows from the set of parent variables to each node, and leave edges blank between nodes that are conditionally independent under this assignment of minimal parents.

**Definition 1.15.** If a probability  $p$  admits a factorisation of the form (1.14) relative to a DAG  $G$  we say that  $G$  represents  $p$ , or that  $G$  and  $p$  are compatible, or that  $p$  is Markov relative to  $G$ .

A helpful criterion that we will appeal to later is the graphical notion of d-separation. This notion is useful for determining if a distribution is compatible with a given DAG, by a set of conditional independencies that must be satisfied by the distribution itself.

**Definition 1.16.** A path is said to be d-separated by a set of nodes  $Z$  if and only if:

1. the path contains a chain  $X_i \rightarrow X_m \rightarrow X_j$  or fork  $X_i \leftarrow X_m \rightarrow X_j$  such that  $X_m$  is in  $Z$ .
2. the path contains a collider  $X_i \rightarrow X_m \leftarrow X_j$  such that  $X_m$ , or no child of  $X_m$  is not in  $Z$ .

A set  $Z$  is said to d-separate  $X$  from  $Y$  if and only if  $Z$  d-separates every directed path from  $X$  to  $Y$ . If  $X$  and  $Y$  are d-separated by a set of nodes  $Z$ , we write  $(X \perp\!\!\!\perp Y | Z)_G$ .

There is a strong association between d-separation and probabilistic conditional independence. If sets  $X$  and  $Y$  are d-separated in a DAG  $G$ , then  $X \perp\!\!\!\perp Y|Z$  in every distribution that is compatible with the DAG  $G$ . Conversely, if  $X$  and  $Y$  are not d-separated by  $Z$  in  $G$ , then there is at least one distribution, compatible with  $G$  where  $X$  and  $Y$  are conditionally dependent. The Markovian parental condition has some strong implications for the particular DAG structure that can be formed. Namely, a necessary and sufficient condition for a probability distribution  $p$  to be Markov relative to a DAG  $G$  is that every variable be conditionally independent of all its non-descendants in  $G$ , conditional on its parents. Some authors call this criterion the ‘local’ Markov condition and it is often taken as the definition of a Bayesian network.

Note that there can be a variety of probability distributions that are compatible with a DAG  $G$ , so as to say the assignment from DAG’s to probabilities is not unique. We say that two DAG’s are observationally equivalent if and only if they have the same skeletons and the same sets of collider structures. This places a limit on our ability to infer the direction of edge assignments in the DAG structure by observing probabilities alone, however, as we will see in the next sections, there may be an experiment we can perform on the process that generates these probabilities that will allow us to assign them correctly.

We now will discuss how to place structure onto a DAG  $G$ , removing the arbitrariness of just some probability distribution. In generality, a structural model is a finite set of equations over a finite number of variables of the form

$$(1.48) \quad x_i = f_i(pa_i, u_i), \quad i = 1, \dots, n$$

where as before  $pa_i$  denotes the set of variables that directly determine the value of  $X_i$ , and  $u_i$  represents a noise term that is in the classical case assumed to be

1.  $u_i \perp\!\!\!\perp u_j | \emptyset$ , which is to say, each error term is independently distributed and so not correlated and,
2. Arising due to ignorance of latent, or unobserved variables.

A simple example of structural models are linear structural models, which are often referred to as linear structural equation models and are of the form

$$(1.49) \quad x_i = \sum_{k \neq i} \alpha_{ik} x_k + u_i, \quad i = 1, \dots, n$$

and have become commonplace among a variety of fields, such as economics, social science and machine learning. The interpretation of the causal model and the mechanistic structure of the functions relating variables will of course be different across all applications, however this basic functional form suffices to tell us answers to important questions we might have in any causal model, such as how does the variable  $y$  change when I tinker with variable  $x$ ?

**Definition 1.17.** *A causal model is a pair  $M = \langle D, \Theta \rangle$  consisting of a causal structure  $D$  and a set of parameters compatible with  $D$ . A causal structure of a set of variable  $V$  is a DAG in which each node corresponds to distinct elements of  $V$  and each edge represents a direct functional (as in, not through or via other variables in the graph) relationship among the linked variables. the parameters  $\Theta$  assign a function  $x_i = f_i(pa_i, u_i)$  for each variable in the model, and a probability measure  $p(u_i)$  to each disturbance  $u_i$ , which is independent across all the other  $u_{j \neq i}$ .*

Probabilities over classical variables then arise naturally as the result of the probability distributions that unobserved variables in the model are assumed to take. However, we shall see that there is no need to introduce background unknowns into causal models to introduce probability. It is possible to use the intrinsic properties of quantum theory to do so. One might imagine that structural models are also useful tools for describing quantum experiments, where the function  $f_i$  is the Born Rule. We will show that this is indeed possible and interesting to investigate. Before we reach this departure, we make some remarks on refinements to this definition.

**Definition 1.18.** *A structural causal model is a triple  $M = \langle U, V, F \rangle$  where*

1.  *$U$  is a set of unobserved variables, also referred to as exogenous that are determined by factors outside of the model  $M$ .*
2.  *$V$  is a set of observed variables, also referred to as endogenous that are determined by variables in the model.*
3.  *$F$  is a set of functions such that each  $f_i$  is a mapping from  $U_i \cup PA_i$  to  $V_i$ .*

Note the differences between definitions (1.17) and (1.18). In (1.17) the background variables are all assumed to be uncorrelated with one another, whereas (1.18) makes no such assumptions. These two formalisms are readily combined into the structure of probabilistic causal models.

**Definition 1.19.** *A probabilistic causal model is a pair  $\langle M, p(U) \rangle$  where  $M$  is a structural causal model and  $p(U)$  is a probability distribution defined over the background variables.*

We will now make some claims that bring the concept of time into the paradigm of causal modelling. Determining the causal structure of data in the absence of temporal information brings about some interesting philosophical possibilities. In the absence of temporal information, and since independence is a purely statistical phenomena, do we ever reach assignments of causal links in a DAG that lead to a clash with temporal knowledge? For instance, is it ever possible to assign the edge  $X \rightarrow Y$  when temporal data would have us conclude that the correct relation is given by  $Y \rightarrow X$ ? Being a statistical phenomena only, it is indeed possible to manufacture data that behave this way.

The temporal aspect of data serves us in the following way. Effects should precede causes. The statistical aspect however is only there to render complete causal explanations of data through

conditional independences which screen influence. Statistical explanations of phenomena or data without such causal influence are deemed ‘incomplete’ and the residual dependencies are considered ‘spurious’ or ‘unexplained’. The fact that the temporal and causal descriptions of data have existed parsimoniously throughout centuries of scientific experiment and, suffice to say, human experience has led to the formation of the following ideas. The first is of statistical time.

**Definition 1.20.** *Given an empirical distribution  $p$ , a statistical time of  $p$  is any ordering of the variables that agrees with at least one minimal causal structure agreeing with, or Markov relative to  $p$ .*

As there may be many DAG structures that are simultaneously compatible with data observed in an experiment, some model selection criteria must be used. There are a variety of ways to score a causal model but the simplest would be an Occam’s razor type argument. We can formalise this thus. The DAG models in any probabilistic causal model compatible with  $p$  necessarily are equipped with a set of probabilities and latent variable structures. Application of Occam’s razor would lead one to assert that the ‘correct’ causal model is the one that has the simplest structure of probabilities and latent variables, while still being compatible with  $p$ .

However, when invoking temporal information we also have natural law to resort to. We should then pick the model according to the simplest model within the class of models which have a statistical time that is compatible with physical time. In fact, there remains an open conjecture [58] here that we state for completeness.

**Postulate 7.** *In all natural phenomena, the physical time coincides with at least one statistical time.*

### 1.2.3 Levels of the Pearl Hierarchy

There have been many recent developments concerning causal structures and how to account for differing modes of reasoning in a causal model. What is a causal model and how is it useful? Pearl believes the answer lies in a three stage hierarchical structure of predicting, intervening, and reasoning. A prediction task is one where we should like to estimate the probability that our observational data will be of a certain value, given that we have observed its causal parents to be at some value. A typical activity where prediction is essential is seeing, and asking ‘what is?’ type questions. The association is to probabilities of the type  $p(y|x)$  and makes statements about the rational way I should change my beliefs about a system  $Y$  given that I see  $X$  in some state.

For instance, consider the causal diagram in figure 1.5. We might ask ourselves, given that I have observed the barometer needle at a certain position, what is the probability that there will be a thunderstorm.

Next in the Pearl hierarchy comes intervening, which is the process of an observer performing an intervention on some system. The probabilities that represent degrees of belief in this paradigm are often denoted  $p(y|\text{do}(x))$ , and can be quite different from the associational probabilities.



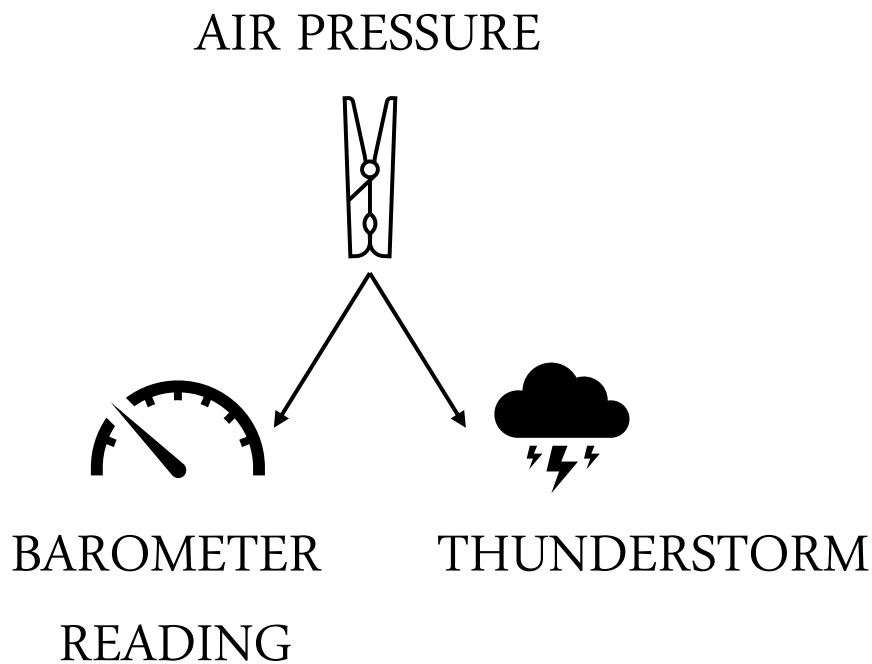


Figure 1.5: A simple causal diagram relating the air pressure, a reading on a barometer needle, and whether the weather is a thunderstorm.

For example, if I once again observe the barometer needle, air pressure and thunderstorm causal system, then the probability of there being a thunderstorm is unaltered if I artificially interact with the barometer needle, and set it to some particular value  $do(x)$ . In fact, doing so does not alter the probability for me to observe a thunderstorm since there is no direct causal link between thunderstorm and barometer needle. They are both descendants, or proxies for the air pressure and by altering the position of the needle on the barometer, there is no alteration in air pressure and so no change observed in the frequency of thunderstorms. The types of queries the international probabilities allow us to answer are ‘what if?’ type questions, which are radically different queries to ‘what is?’.

The third layer of the Pearl hierarchy is concerned with counterfactual reasoning. Counterfactual reasoning is a process by which we imagine how things would be different in a world that did not exist. Counterfactual queries cover cases such as ‘what if I had acted differently’. The nomenclature of a counterfactual probability is  $p(y_x|x, y')$  which would be read as ‘the probability I would have observed an effect  $y_x$  when I observed the cause to be  $x$ , when in fact and in the real world I saw the effect to be  $y'$  and the cause to be  $x'$ . Was it really the aspirin that cured my headache is asking would I still have a headache if I did not take the aspirin, the two worlds cannot exist simultaneously but we can reason about them none-the-less. In fact, a better word than reason is to perhaps imagine, and there is a growing body of literature pointing to the idea that endowing machines with this type of reasoning is essential to produce a true artificial intelligence. This type of reasoning is something that we become comfortable with as children,

but despite this, ideas for formalising counterfactual probabilistic reasoning are very much a late 20th and 21st century development.

Moving between the layers of the Pearl hierarchy is a combination of utilising the underlying causal model for the process that generates the data we wish to ask questions of, and three rules that combine together to form the do-calculus.

**Definition 1.21.** *let  $X, Y, Z$  be disjoint sets of nodes in a DAG  $G$ . we denote by  $G_{\bar{X}}$  the graph obtained by deleting all arrows in  $G$  that point to nodes in  $X$ . We denote by  $G_{\underline{X}}$  the graph obtained by deleting all arrows from  $G$  emerging from nodes in  $X$ . We denote  $G_{\bar{X}\underline{Z}}$  the graph obtained by deleting arrows into  $X$  and out of  $Z$ . Finally we denote  $p(y|do(x), z) = p(y|\hat{x}, z) = p(y, z|\hat{x})/p(y|\hat{x})$  the probability that  $y = Y$  given that  $X$  is held constant at  $x$  and  $Z = z$  is observed. The rules of the do-calculus associated to a causal model are then given as:*

1.  $p(y|\hat{x}, z) = p(y|\hat{x}, z, w)$  if  $(Y \perp\!\!\!\perp Z|X, W)_{G_{\bar{X}}}$ .
2.  $p(y|\hat{x}, \hat{z}, w) = p(y|\hat{x}, z, w)$  if  $(Y \perp\!\!\!\perp Z|X, W)_{G_{\bar{X}\underline{Z}}}$ .
3.  $p(y|\hat{x}, \hat{z}, w) = p(y|\hat{x}, w)$  if  $(Y \perp\!\!\!\perp Z|X, W)_{G_{\bar{X}\underline{Z}(W)}}$  where  $Z(W)$  is the set of  $Z$  nodes that are not parents or ancestral parents of any  $W$  node in  $G_{\bar{X}}$ .

Each of these inference rules has its origin in the intuition that the action of intervening in a causal model with  $do(x)$  is to ignore any previous dependency that  $X$  had on its ancestors and sever the links, then substitute a new value in as the value for  $X$ . This intuition is sufficiently powerful to explain why wiggling barometer needles is not a useful thing to do should you wish to predict the weather, since you are cutting the causal link from air pressure to barometer needle position. The first rule says that it is valid to delete observations from the data that are independent of the variable we wish to predict, and affirms the connection between the graphical notion of d-separation and the statistical notion of conditional independence. The second is a condition that actions at some node have the same effect as observations at a different node. The third provides a condition for introducing or deleting external interventions without effecting the probability of the target variable.

The effect of acting in a causal model can be predicted if there exists a transformation using the do-calculus rules of inference from a probability involving the hatted variables to an expression containing unhatted variables. In such cases, we say the query is identifiable. Hopefully, any and all uses of the hat operator will not contradict themselves between the usage in quantum theory for the operator associated to the observable  $X$ , and the do action effect in a causal model. Again, I endeavour to make this clear should it crop up.

### 1.2.4 Quantum Causal Models

An important primer, which we find important to consider should we wish to combine the formalism of DAG's and quantum theory together in a single object is the notion of common cause.

Reichenbach gave a formal statement in [59] that was published posthumously in 1959. Suppose that two events are correlated, and neither is the cause of the other. Reichenbach claims then that the events will have a common cause. What the Reichenbach principle asserts is the blanket statement ‘correlation does not imply causation’ should be amended to ‘no correlation without causation’. We devote this section of the introduction to this result, since it is one of the first avenues that allows us to translate between the familiar classical theory, and the quantum.

The authors of [60] break the formal Reichenbach statement into two parts. First, the qualitative part which says if two variables  $Y$  and  $Z$  are found to be statistically dependant, then either

1.  $Y$  is a cause of  $Z$ .
2.  $Z$  is a cause of  $Y$ .
3. There is no causal link between  $Y$  and  $Z$ , but there is a common cause  $X$  influencing both.
4.  $Y$  is a cause of  $Z$  and there is a common cause  $X$  influencing both.
5.  $Z$  is a cause of  $Y$  and there is a common cause  $X$  influencing both.

where causal links can be either direct (with respect to some DAG  $G$ , the shortest path between the effect and cause is length 1) or indirect (mediated by some other variables). The quantitative part concerns the third case, and is actually a restatement of definition (1.13): if  $X$  is a complete common cause of the variables  $Y$  and  $Z$ , then the joint distribution  $p(x, y, z)$  factorises to  $p(y|x)p(z|x)p(x)$ . In order to obtain a notion of a complete common cause for quantum theory, and so a quantum Reichenbach principle, first the case is made for the link between the qualitative and quantitative parts of the classical Reichenbach principle.

For a classical conditional  $p(Y|X)$ , a classical dilation is a deterministic function  $f(X, \lambda)$  and a probability distribution over  $\lambda$ ,  $p(\lambda)$  such that

$$(1.50) \quad p(Y|X) = \sum_{\lambda} \delta_{Y, f(X, \lambda)} p(\lambda),$$

with  $\delta_{Y, f(X, \lambda)} = 1$  if  $Y = f(X, \lambda)$ . Such dilations appear in the literature, and are however not to be confused with the ‘instrumental variables’ used in the modelling of interventions in a causal model [58]. In that instance, the  $\lambda$  are observed, and more importantly, controlled. The idea is that if the observed  $X$  really is a cause of  $Y$ , and the dynamics are fundamentally deterministic, then any change in  $Y$  we observe for a given value of  $X$  is due to exogenous variations that we have not observed.

Application of this idea to the scenario where  $X$  is a complete common cause of  $Y$  and  $Z$ , that is, there are other causes of  $Y$  and  $Z$  but  $X$  is the only shared cause leads to

$$(1.51) \quad P(YZ|X) = \sum_{\lambda_y, \lambda_z} \delta_{Y, f_Y(X, \lambda_y)} \delta_{Z, f_Z(X, \lambda_z)} p(\lambda_y, \lambda_z)$$

and since we demand, from the causal structure, that  $p(\lambda_y, \lambda_z) = p(\lambda_y)p(\lambda_z)$ , it then follows that  $P(YZ|X) = P(Y|X)P(Z|X)$ . The probability  $P(YZ|X)$  is said to be compatible with  $X$  being the complete common cause if one can find functions  $p(\lambda_y), p(\lambda_z)$  that satisfy equation (1.51). One should note that the causal explanation invoked here is not unique, and by tuning the causal dependencies it is often possible to find more complex models in which  $X$  is not the complete common cause. However, one often rejects these in favour of simpler models and as such the best explanation of  $P(YZ|X) = P(Y|X)P(Z|X)$  is that  $X$  is the complete common cause of  $Y$  and  $Z$ .

In order to find a quantum mechanical generalisation of Pearliean causal models, a good place to start is by generalising the Reichenbach statement. In fact, the qualitative part of the Reichenbach statement is trivial to generalise. One can take the conditions listed above and replace all classical variables  $X, Y, Z$  with labels for quantum systems  $A, B, C$  and replace the words ‘statistically dependent’ with the word ‘correlated’, where here  $A, B$  correlated means the joint state over  $A$  and  $B$  does not factorise over the tensor product.

The quantitative part however, is more subtle. Defining an analogue of  $P(YZ|X)$  can be achieved by appealing to the formalism of quantum channels and their unitary dilation.

**Definition 1.22.** For a quantum channel  $\mathcal{E}_{B|A}$  a unitary dilation is given by some ancillary system  $\lambda$  with state  $\rho_\lambda$  and a unitary  $U$  from  $\mathcal{H}_A \otimes \mathcal{H}_\lambda$  to  $\mathcal{H}_B \otimes \mathcal{H}_{B'}$  such that

$$(1.52) \quad \mathcal{E}_{B|A}(\cdot) = \text{Tr}_{B'}(U(\cdot \otimes \rho_\lambda)U^\dagger)$$

where  $\dim(\mathcal{H}_{B'})$  must satisfy  $d_A d_\lambda = d_B d_{B'}$ .

An alternative way to describe channels is through the Choi–Jamiolkowski isomorphism, where states and channels share a duality through

$$(1.53) \quad \rho_{B|A} = \sum_{i,j} \mathcal{E}_{B|A}(|i\rangle\langle j|) \otimes |i\rangle\langle j|.$$

Suppose that the state is  $\rho_B = \mathcal{E}_{B|A}(\rho_A)$ . We can also describe this in terms of the state representation of the channel, via the linking operator

$$(1.54) \quad \tau_A = \sum_{l,m} |l\rangle\langle m| \otimes |l\rangle\langle m|$$

such that  $\rho_B = \text{Tr}_A(\rho_{B|A}\tau_A\rho_A)$ . Finally we define the notion on no-influence. Under a unitary channel  $\rho_{BB'|AA'}$ ,  $A$  has no causal influence on  $B$  if and only if for  $\rho_{B|AA'} = \text{Tr}_{B'}(\rho_{BB'|AA'})$  we have  $\rho_{B|AA'} = \mathbb{1}_A \otimes \rho_{B|A'}$ . This can be rephrased as  $A$  having no causal influence over  $B$  if and only if any local operations carried out on the state before it enters the channel at  $A$  do not effect the state at  $B$ . The generalisation of the classical notion of compatibility with a complete common cause then is stated as follows:

**Definition 1.23.**  $\rho_{BC|A}$  is compatible with  $A$  being the complete common cause of  $B$  and  $C$  if it is possible to find ancillary systems  $\lambda_B, \lambda_C$  and states  $\rho_{\lambda_B}, \rho_{\lambda_A}$  and a unitary channel where  $\lambda_B$

has no influence on  $C$  and simultaneously  $\lambda_C$  has no influence over  $B$ , such that  $\lambda_B, \lambda_C$  admit a dilation of  $\rho_{BC|A}$ .

A quantum generalisation of Reichenbach's principle is then given by the equivalence between  $\rho_{BC|A}$  being compatible with  $A$  being the complete common cause of  $B$  and  $C$ , and the statement that  $\rho_{BC|A} = \rho_{B|A}\rho_{C|A}$ , where identities are suppressed as required.

This brings us to a definition of a quantum causal model. Note that this is an ongoing area of research, and definitions may vary between authors.

**Definition 1.24.** *A quantum causal model is given by*

1. *A DAG  $G$ , with vertices  $A_1, \dots, A_N$  corresponding to quantum nodes (which is a shorthand for saying the tensor product of input and output Hilbert spaces).*
2. *For each  $A_i$ , a quantum channel  $\rho_{A_i|PA(A_i)}$ , mapping from the output spaces of the parent set of nodes to the output set of  $A_i$ , such that  $[\rho_{A_i|PA(A_i)}, \rho_{A_j|PA(A_j)}] = 0$  for all  $i, j$ .*

This definition comes with its own Markov condition, much like the classical counterpart.

**Definition 1.25.** *Given a DAG  $G$ , with quantum nodes  $A_1, \dots, A_n$  - and a process operator  $\sigma_{A_1, \dots, A_n}$ , we say the process is Markov relative to  $G$  if and only if there exists for each  $i$  a quantum channel  $\rho_{A_i|PA(A_i)}$ , such that  $[\rho_{A_i|PA(A_i)}, \rho_{A_j|PA(A_j)}] = 0$  for all  $i, j$ , and  $\sigma_{A_1, \dots, A_n} = \prod_i \rho_{A_i|PA(A_i)}$ . We define a process operator to be a positive semi-definite operator for which  $\text{Tr}[\sigma_{A_1, \dots, A_n} \tau_{A_1} \otimes \dots \otimes \tau_{A_n}] = 1$  for any set of channels  $\tau_{A_1}, \dots, \tau_{A_n}$  on the nodes.*

As far as we know, this definition or ones very similar to it represent all quantum theory can do in a causal scenario. We will not use these causal models much in this thesis. When we start to prove stuff, we will appeal to our own notation and ideas. However, I wanted to include them because they provide a complete picture of modern causality literature and very nicely lead up a natural question we can ask about causality. What kinds of correlations are outside of a causal scenario - that is, have no explanation in terms of a causal model?

### 1.2.5 Causal Inequalities

Our experience of the conjecture that every statistical time co-insides with at least one physical time is so well founded it seems radical to attempt to diverge from it. It is certainly our common understanding of the world that events which happen are embedded within a causal structure, even if the myriad of confusing and hard to interpret data, as well as many, many unobserved confounding variables, prevents us from discerning this causal structure exactly. But is this a physical principle? Namely, is causal influence always unidirectional in any valid physical theory of nature? We can imagine lifting the acyclicity requirement of DAG structures [61] and be left with a simple directed graph that describes the causal structure in some theory. Does nature ever do this, and what are the ramifications if it did?

Let us begin by defining a causal correlation. We take the bipartite case for simplicity. Imagine two parties, denoted  $A$  for Alice and  $B$  for Bob. Alice and Bob will conduct an experiment whereby they are given some random classical input variables  $x$  and  $y$  respectively and are asked to produce outputs  $a$  and  $b$ . The joint conditional probability produced in this experiment is denoted  $p(a, b|x, y)$ . Consider the situation where Alice's events precede Bob's. We will always for this argument assume that we can well define who comes first and that there is no temporal overlap between the two parties.

In this scenario, Alice could send her output to Bob, but not vice versa and so there is a No-signalling type constraint from Bob to Alice. In this case, we write that

$$(1.55) \quad p_A^{A \rightarrow B}(a|x, y) = p_A^{A \rightarrow B}(a|x, y')$$

for all  $x, y, y'$  where  $p_A^{A \rightarrow B}(a|x, y) = \sum_b p^{A \rightarrow B}(a, b|x, y)$ . Likewise, if Bob is first in the temporal and causal order we have the constraint that

$$(1.56) \quad p_B^{B \rightarrow A}(b|x, y) = p_B^{B \rightarrow A}(b|x', y)$$

for all  $x, x', y$  and where  $p_B^{B \rightarrow A}(b|x, y) = \sum_a p^{B \rightarrow A}(a, b|x, y)$ . A bipartite probability is then deemed causal if and only if it can be written in the form

$$(1.57) \quad p(a, b|x, y) = qp^{A \rightarrow B}(a, b|x, y) + (1 - q)p^{B \rightarrow A}(a, b|x, y).$$

Being a probability,  $q \in [0, 1]$ , so that this corresponds to picking the causal order in a particular realisation of the experiment from a probability distribution. Note that by randomly using different causal strategies in an experiment, we always use a causal strategy *overall*. This property is called the convexity of the set of causal correlations. Since the restrictions on the probabilities are linear, this convex set is also a polytope. We denote the set of correlations which can be written in the form (1.57) the causal polytope. Being probabilities, there are a variety of restrictions on the polytope that are not causal in nature. These are the non negativity  $p^{A \rightarrow B}(a, b|x, y) > 0, p^{B \rightarrow A}(a, b|x, y) > 0$  constraints and the normalisation  $\sum_{a,b} p^{A \rightarrow B}(a|x, y) = \sum_{a,b} p^{B \rightarrow A}(a|x, y) = 1$  constraints. The other, non-trivial facets of this polytope are called causal inequalities.

Any physical distribution generated in accordance with a causal process will satisfy causal inequalities. Before we digress any further into the subject of causal inequalities, we will finish our characterisation of the bipartite two-input two-output causal polytope. The vertices of the bipartite causal polytope can be seen to correspond to deterministic strategies for generating probabilities, which we will now show. Evident from an application of Bayes rule and the Markovian parental condition we have

$$(1.58) \quad p^{A \rightarrow B}(a, b|x, y) = p(a|x)p(b|x, y, a) = p(a|x, y)p(b|x, y, a).$$

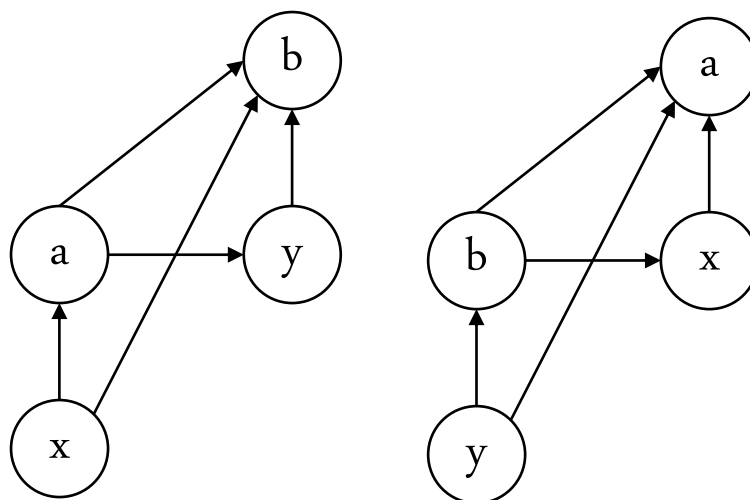


Figure 1.6: The two permissible DAGs for causal models in the causal scenario described. These cover all possible probabilities consistent with the temporal ordering  $t_x < t_a < t_y < t_b$  on the left and  $t_y < t_b < t_x < t_a$  on the right, with independent inputs. If a bipartite probability distribution cannot be explained as arising from a mixture of these causal models, and there are no latent variables in the model (i.e. the system is well isolated) then we say that the correlation is not causal, and violates a causal inequality. All causal correlations for two parties can be written as in equation (1.57).

By writing these probabilities as convex functions of deterministic functions  $\alpha(x)$  and  $\beta(x, y, a)$  of the input variables

$$(1.59) \quad \begin{aligned} p(a|x) &= \sum_{\alpha} q_{\alpha} \delta_{a, \alpha(x)} \\ p(b|x, y, a) &= \sum_{\beta} q_{\beta} \delta_{b, \beta(x, y, a)} \end{aligned}$$

with  $q_{\alpha}, q_{\beta} > 0$  and  $\sum_{\alpha} q_{\alpha} = \sum_{\beta} q_{\beta} = 1$ . We may then write (1.58) as

$$(1.60) \quad p^{A \rightarrow B}(a, b|x, y) = \sum_{\alpha, \beta} q_{\alpha} q_{\beta} \delta_{a, \alpha(x)} \delta_{b, \beta(x, y, a)},$$

and hence any  $p^{A \rightarrow B}$  can be written as a convex combination of deterministic probability distributions. This in turn implies the deterministic probability distributions sit at the vertices of the causal polytope. In the case of the two-input, two-output and two-party scenario, there are 112 such vertices. Any convex polytope has an equivalent representation in terms of its sides, or facets, and it is important to obtain both representations for any causal polytope since in this facet representation we find our causal inequalities. The authors of [62] enumerated the facets of the two-input, two-output bipartite polytope and found that there are 48 such facets.

16 of these are trivial constraints on the non-negativity of  $p(a, b|x, y)$ . The remaining 32 fit within two separate classes, which can be physically motivated. The first class contains 16 facets

and are all relabellings of the inequality

$$(1.61) \quad \frac{1}{4} \sum_{x,y,a,b} \delta_{a,y} \delta_{b,x} p(a,b|x,y) \leq \frac{1}{2}$$

and the remaining 16 are relabelling of the inequality

$$(1.62) \quad \frac{1}{4} \sum_{x,y,a,b} \delta_{x(a \oplus y),0} \delta_{y(b \oplus x),0} p(a,b|x,y) \leq \frac{3}{4},$$

where  $\oplus$  denotes binary addition.

We can physically motivate the inequality (1.61). Consider the players Alice and Bob in the two-input, two-output schema, who are tasked with guessing the input bit of the other. Assume that the input bits are distributed randomly with equal probability. They are permitted to communicate, but before Alice can communicate anything to Bob she must produce her output variable and before Bob can communicate to Alice he must produce his output. Without this restriction on communication, the game becomes trivial since it is possible to generate  $p(a,b|x,y)$  with unlimited rounds of communication, in some sense rendering the game local. By rewriting the inequality (1.61) as

$$(1.63) \quad p(a=y, b=x) \leq \frac{1}{2},$$

this causal inequality represents the following scenario. In the causal order  $A \rightarrow B$  ( $B \rightarrow A$ ), Alice (Bob) has no knowledge of Bob's (Alice's) input, and so can only make a random guess at what his (her) input is. She then communicates her (his) input to Bob (Alice) who can now with certainty get the correct answer for Alice's (Bob's) input. Hence, the total success probability in this game is bounded by how well Alice can guess a random uniformly drawn bit, which is  $1/2$ . This game has come to be known as the guess your neighbours input game [63] and represents one of the simplest causal inequalities known.

Similarly, (1.62) can be motivated by a lazy version of the same game, wherein the players are tasked to guess only when their respective input is 1, and always win when their input bit is 0.

Importantly, it has been shown that [63] quantum theory does not violate these inequalities. This, and generalisations thereof will be a key topic in this thesis, especially in Chapter 3.

Comparisons can be drawn between the causal polytope described and the local polytope from the theory on No-signalling. The local polytope arises through much a similar construction. Recall that a local correlation can be written in the form (1.24), which, for the two-input-two-output bipartite case can be written in the form

$$(1.64) \quad p(a,b|x,y) = \sum_{\lambda} p(\lambda) p(a|\lambda,x) p(b|\lambda,y)$$

where  $\lambda$  represents some shared random variable and  $p(\lambda)$  the exogenous probability distribution that  $\lambda$  occurs with. Again, deterministic probabilities sit at the vertices of the polytope of local



correlations, and for the two-input, two-output case can be straightforwardly enumerated. We can write these as

$$(1.65) \quad p(a, b|x, y) = \begin{cases} 1 & a = \alpha x \oplus \beta, b = \gamma y \oplus \delta \\ 0 & \text{else} \end{cases}$$

where  $\alpha, \beta, \gamma, \delta \in \{0, 1\}$  are the four parameters that take on all possible combinations of zeros and ones to give 16 possible vertices. By utilising the tools of linear programming, we are able to convert this vertex representation into an facet representation, giving the non-trivial facets of the local polytope as

$$(1.66) \quad (-1)^\gamma E(0, 0) + (-1)^{\beta+\gamma} E(0, 1) + (-1)^{\alpha+\gamma} E(1, 0) + (-1)^{\alpha+\beta+\gamma+1} E(1, 1) \leq 2$$

which are relabellings of the CHSH inequality (1.30). Hence, causal inequalities and Bell inequalities share a correspondence in classical physics as being the hyper-planes separating achievable correlations for non-achievable correlations. For the case of the local polytope, we know that Bell inequalities are violated if the experimenters are permitted to use entangled quantum states, which begs the question is quantum theory able to violate causal inequalities as well? It was known much prior to this work that the causal inequalities (1.61) and (1.62) are not violated by quantum theory. But there is still the option that a more exotic causal inequality could be violated by quantum theory.

We have considered the bi-partied case for simplicity. In-fact, some work that I [64] and others [65] have done show that, much like the theory of non-locality, there are phenomenological differences between causal inequalities, that make themselves apparent once we move out of the restrictive, two party, two output, two input scenario. More on this, in chapter 4.

## 1.3 Post-Selection

In probability theory, to post-select is an act of conditioning the data upon the occurrence of a given event. Seen as a simple tossing away of data, we might wonder if this is ever a useful thing to do. We hope to motivate this concept sufficiently that the reader will believe that it is.

### 1.3.1 Connections to Quantum Mechanics

The standard procedure for statistical inference in quantum mechanics is to input an initial state into some experimental apparatus and perform measurements upon it. This procedure can represent, abstractly, quantum experiments and display quantum effects such as interference and entanglement. It is not a restrictive paradigm, and I do not wish to claim that it is. However, there is another pedagogy one can employ when conducting experiments [66]. It is also a valid thing to do to specify, independently of the initial state construction, the final state of a system. A simple thought experiment to demonstrate the concept is this: at some initial time  $t_1$  an experimenter,

Alice, prepares a quantum state  $|\phi\rangle$ . In the time interval between  $t_1$  and the final time of the experiment  $t_2$  she performs some arbitrary experiments and records the results. Then, at  $t_2$ , she measures an observable  $A$ , with non-degenerate eigenvalues  $|\psi\rangle$ . If she obtains the outcome associated to  $|\psi\rangle$  in her measurement, she is sure that the state of the system at time  $t_2$  is indeed  $|\psi\rangle$  and keeps all of her statistics. If instead she obtains a different outcome then she is sure that the state of the system is not  $|\psi\rangle$  at time  $t_2$  and discards all of her statistics. In this way, by repeating the procedure many times, Alice has constructed a pre-and-post-selected ensemble of states  $|\phi\rangle$  at  $t_1$  and  $|\psi\rangle$  at  $t_2$ .

The important thing to notice is that the statistics obtained in any intermediary measurement performed for  $t$  such that  $t_1 < t < t_2$  can be quite different from the case of non-post-selected quantum theory. It often cannot be known *ex ante* whether the post-selection upon the ensemble will succeed with certainty, much in the same way that outcomes for measurements cannot be known in regular quantum theory ahead of time, Alice can only make the post-selection after the fact.

Consider the case where the intermediate measurement is detailed, which is defined as the case where each eigenvalue corresponds to one POVM element of the measurement, no information is left unread and as such detailed POVM's evolve pure states to pure states<sup>3</sup>. This implies that the normalisation for POVM elements  $E_a$  can be written as  $\sum_a E_a^\dagger E_a = \mathbb{1}$ . The probability to obtain outcome  $a$  in the pre-and-post-selected ensemble is then given by the following formula.

**Definition 1.26.** *Given a pre and post-selected quantum state, where  $|\phi\rangle$  corresponds to the pre-selection and  $|\psi\rangle$  corresponds to the state which passes the post-selection, and a detailed measurement made at the intermediary times well described by the POVM elements  $E_a$ , the probability to obtain outcome  $a$  can be calculated and given as*

$$(1.67) \quad p(a) = \frac{|\langle\psi|E_a|\phi\rangle|^2}{\sum_{a'} |\langle\psi|E_{a'}|\phi\rangle|^2}.$$

This rule is known as the ABL formula, names for its developers Aharonov, Bergmann and Lebowitz. The ABL rule can easily be generalised to the case where the pre-selection and post-selection are on density matrices and not pure quantum states, as we shall see in chapter 3. The case where the input state is pure, but the post-selection can be passed for a multiplicity of degenerate eigenvalues of an observable  $B = b$  is given by

$$(1.68) \quad p(a) = \frac{\sum_b |E_b E_a |\phi\rangle|^2}{\sum_{a',b'} |E_{b'} E_{a'} |\phi\rangle|^2}$$

which allows the connection to regular quantum theory to be made explicit. If there is no post-selection, we are never in the case where we discard statistics. the measurement that returns the

<sup>3</sup>for the more general case, where the state can remain entangled under a POVM see for example [66]

'keep statistics' answer for any state is exactly a measurement of the identity operator  $\mathbb{1}$ , which substitution into this form of the ABL formula gives

$$(1.69) \quad p(a) = |E_a |\phi\rangle|^2$$

which is postulate four of quantum mechanics.

The most famous use for the two-time states is the weak measurement. A weak measurement is a measuring procedure, described by an interaction Hamiltonian where the coupling constant is typically small (although this is by no means a fundamental restriction upon the framework). In a typical measurement, we place ourselves in the regime where the device we will use to readout the measurement is well localised around 0 in position. We do this so that the value, instead of the change of value of the pointer state will be proportional to the measured quantity. However, well localised around the 0 state means that the pointer will have a large uncertainty in its momentum and as such a large uncertainty in the Hamiltonian implementing the correct measurement. In a weak measurement, the pointer on the measuring device is well localised around the zero in momentum and the position reading of the pointer becomes imprecise. It is only by performing the weak measurement many times (for a system of  $N$  particles we improve the resolution on our device by a factor of  $\sqrt{N}$ ) we obtain accurate results. The outcome of weak measurements are dubbed weak values, let us now show how these weak values arise from pre-and-post-selected ensembles. The standard implementation of a Von-Neumann measurement of an observable  $A$  is through an interaction Hamiltonian of the form

$$(1.70) \quad H_{int} = g(t)\hat{p} \otimes \hat{A}$$

The coupling  $g(t)$  need not be small by the way, it is common to impose the condition

$$(1.71) \quad \int dt g(t) = 1.$$

A pointer on a measuring device initially in the Gaussian state with standard deviation in  $x$  of  $\Delta$  is:

$$(1.72) \quad |\phi^{MD}\rangle = \int dx (\Delta\pi)^{-1/4} e^{-x^2/(2\Delta^2)} |x\rangle$$

can be shown to move to a Gaussian centred about the expectation value of the measurement  $\langle\phi|A|\phi\rangle$ . We say that the weak value in a standard (non-post-selected) measurement agrees with the expectation value, however they arise due to quite conceptually different objects, one the weighted average of eigenvalues associated to the operator  $A$  and the other a physical shift in a pointer state of which taking averages gives the expectation value.

For a pre and post-selected ensemble, the process looks similar, but with some notable differences [55]. We assume the free Hamiltonian is zero, and that our system is pre-selected in the state  $|\phi\rangle$  at  $t_1$  and post-selected in the state  $|\psi\rangle$  at  $t_2$ . The weak measurement takes place during  $t$  between  $t_1$  and  $t_2$ , so that  $g(t) = 0$  if  $t < t_1$  to  $t > t_2$ . Using the same initial state as before,

up-to normalisation, the state of the measurement device after the measurement is complete is given by

$$(1.73) \quad |\phi^{MD}\rangle = \int dx \langle \psi | e^{-i\hat{p}\hat{A}} | \phi \rangle e^{-x^2/2\Delta^2} |x\rangle$$

by taking the Fourier transform, and defining  $A_w = \langle \psi | A | \phi \rangle / \langle \psi | \phi \rangle$ ,

$$(1.74) \quad \tilde{\phi}^{MD} = \langle \psi | \phi \rangle e^{-iA_w p} e^{-p^2\Delta^2/2} + \langle \psi | \phi \rangle \sum_{n=2}^{\infty} \frac{(ip)^n}{n!} (\langle \psi | A^n | \phi \rangle - A_w^n) e^{-p^2\Delta^2/2}.$$

If the spread in the position state of the pointer is initially sufficiently large, we can ignore the second term. Then, by inverting the Fourier transform we find the state of the pointer to be

$$(1.75) \quad |\phi^{MD}\rangle \approx \int dx (\Delta\pi)^{-1/4} e^{-(x-A_w)^2/(2\Delta^2)} |x\rangle$$

which is a Gaussian state with average at the weak value,  $\langle \psi | A | \phi \rangle / \langle \psi | \phi \rangle$ .

**Definition 1.27.** *The weak value of a quantum observable  $A$  and a pre-and-post-selected ensemble of states is given by*

$$(1.76) \quad A_w = \langle \psi | A | \phi \rangle / \langle \psi | \phi \rangle$$

Weak measurements have been shown to be of important practical applications, where they lead to surprisingly high precision measurements of the properties of systems above and beyond that of their typical counterparts. It is also possible for weak measurements to show exotic behaviours beyond those possible within quantum theory without post-selection, for example, the weak value can lie outside of the support of the range of permitted eigenvalues of a typical measurement of the system, and has lead to many startling results. Aharonov *et al.* have a famous result where one is able to show that a spin measurement of a spin 1/2 particle (which, typically can only take on the values of  $\pm 1/2$ ) can turn out to be 200, or larger [67]!

Weak measurements are able to implement any general quantum measurement - with [68] or importantly without ancilla [69].

It is also noteworthy that the weak value has interpretation even for non post-selected systems. By sequentially applying the weak measurement [70] procedure to a quantum state the authors of [71] are able to observe weak values outside of the observable spectrum, without the need to post-select a final state for the system. They then go on to use this procedure in a challenge where the goal is to distinguish between a set of candidate causal orders for a process. In fact, many quantum measurements could be said to be weak measurements - if one defines the phenomena simply as an irreversible accumulation of information and an orthogonality developing between branches in a superposition simultaneously (as opposed to a strong measurement, where first the branching occurs and is then followed by the irreversible accumulation of information by means of interaction with the observer).

### 1.3.2 Self-Controlled Unitary Gates

Another interesting gadget that can be produced by applying post-selection to a system, this example attributable to the author, is that of a self controlled unitary gate. A controlled unitary gate in the field of quantum computing, where we imagine the dimension of the systems is 2 so that we work with qubits is a unitary of the form

$$(1.77) \quad \hat{V} = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes \hat{U}$$

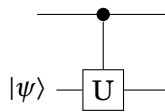


Figure 1.7: Circuit diagram notation for a controlled gate.

which corresponds to the circuit diagram in figure 1.7. The left hand side of the tensor product pertains to a ‘control’ Hilbert space  $\mathcal{H}_C$  and the right hand side pertains to a target  $\mathcal{H}_T$ . Should the control be set to state  $|0\rangle_C$ , the gate is not applied to the target and should the control be set to  $|1\rangle_C$  a unitary  $\hat{U}$  acts upon the target space. Controlled gates form the most primitive objects in quantum computing that are not single qubit gates, all computationally complete gate-sets for quantum computing contain at least one two-or-more qubit gates with the ability to create entanglement, of which controlled gates are perhaps the simplest.

One question that I asked early in my PhD is the following : is it possible to use the state of the target at an earlier time to act as the control for a qubit at a later time? This corresponds to the circuit diagram in figure 1.8.

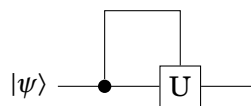


Figure 1.8: Circuit diagram notation for a self-controlled gate.

Does there exist a unitary self-controlled gate in quantum theory? The answer, is no. One can prove this by asserting that the self-controlled operation is not time-reversible, or by having access to one would allow one to do Landauer erasure [72] at no work cost. However, such gates can be realised in post-selected quantum physics. Letting the gate be defined by

$$(1.78) \quad \hat{V} = |0\rangle\langle 0| + \hat{U} |1\rangle\langle 1|$$

where we note that the tensor product structure is dramatically different from (1.77)<sup>4</sup>. Letting  $\hat{U}$

---

<sup>4</sup>although, in the case of the unitary  $\hat{U}$  being diagonal in the basis of the control the gate is trivial to implement, since  $\hat{V}$  is then unitary itself

be the unitary

$$(1.79) \quad \begin{aligned} \hat{U}|0\rangle &= \alpha|0\rangle + \beta|1\rangle \\ \hat{U}|1\rangle &= \gamma|0\rangle + \delta|1\rangle \end{aligned}$$

which gives  $\hat{V} = |0\rangle\langle 0| + \gamma|0\rangle\langle 1| + \delta|1\rangle\langle 1|$ , it is possible to express  $\hat{V}$  (using the singular value decomposition) as a diagonal matrix sandwiched by unitaries,  $\hat{V} = \hat{U}_2 \hat{D} \hat{U}_1^\dagger$ . Inverting the singular value decomposition to make the diagonal matrix the subject gives  $\hat{U}_2^\dagger \hat{V} \hat{U}_1 = \hat{D}$ , where up-to normalisation  $\hat{D} = c(|0\rangle\langle 0| + \epsilon|1\rangle\langle 1|)$ , with  $\epsilon < 1$  if  $\hat{U}$  is not diagonal in the control basis. This diagonal matrix, clearly changes the length of the vectors that it acts upon. So, given access to all single qubit unitaries (and in particular  $\hat{U}_1$  and  $\hat{U}_2^\dagger$ ) and  $\hat{V}$  we can favourably diminish the length of a quantum state vector in a particular direction and stretch it in another, by applying  $\hat{U}_2^\dagger \hat{V} \hat{U}_1$ . Therefore, a single application of  $\hat{V}$  can be seen as one implementing post-selection on the (un-normalised) state  $|\phi\rangle = |0\rangle + \frac{\gamma}{1-\delta}|1\rangle$ . By chaining together a large number of self-controlled gates and the single qubit unitaries we can converge to post-selection upon the  $|0\rangle$  state.

The association is both ways, and as such we now show how we can build one of these self-controlled gates  $\hat{V}$  in quantum theory using post-selection. We make use of an entangled Bell-pair of states over  $\mathcal{H}_1 \otimes \mathcal{H}_2$  and a completely arbitrary state  $|\psi\rangle$  in  $\mathcal{H}_3$  to implement this transformation. First the Bell state is prepared. Then, a standard controlled gate of the form (1.77) is applied with  $\mathcal{H}_3$  as the control and  $\mathcal{H}_1$  as the target. A Bell measurement (i.e. a measurement that returns yes if the particles are in the state  $|\phi^+\rangle$ ) is applied between  $\mathcal{H}_2 \otimes \mathcal{H}_3$ , and post-selected for, and at the end of the process we are left with the state  $\hat{V}|\phi\rangle$  in Hilbert space  $\mathcal{H}_1$ . One can think of this as a combination of the phase kick-back technique common in quantum computing, combined with a post-selection. Formally, this corresponds to the following sequence of steps (with normalisation excluded in this demonstration - see Chapters 4 and 5),

$$(1.80) \quad \begin{aligned} (|0\rangle_1 |0\rangle_2 + |1\rangle_1 |1\rangle_2)(\alpha|0\rangle_3 + \beta|1\rangle_3) &= \alpha|000\rangle_{123} + \alpha|110\rangle_{123} + \beta|001\rangle_{123} + \beta|111\rangle_{123} \\ &\rightarrow \alpha|000\rangle_{123} + \alpha|110\rangle_{123} + \beta\hat{U}|0\rangle_1|01\rangle_{23} + \beta\hat{U}|1\rangle_1|11\rangle_{23} \\ &\rightarrow (\langle 00|^{23} + \langle 11|^{23})(\alpha|000\rangle_{123} + \alpha|110\rangle_{123} + \beta\hat{U}|0\rangle_1|01\rangle_{23} + \beta\hat{U}|1\rangle_1|11\rangle_{23}) \\ &\rightarrow \alpha|0\rangle_1 + \beta\hat{U}|1\rangle_1 = \hat{V}|\psi\rangle_1. \end{aligned}$$

The probability for a successful post-selection goes like

$$(1.81) \quad \begin{aligned} Tr(|\psi\rangle_{123}\langle\psi|_{123}|\phi^+\rangle_{12}\langle\phi^+|_{12}) &= \frac{1}{4} Tr_3((\alpha|0\rangle_3 + \beta U|1\rangle_3)(\alpha^*\langle 0|_3 + \beta^* U^\dagger\langle 1|_3)) \\ &= \frac{1}{4} (|\alpha|^2 + |\beta|^2 + \alpha^*\beta\langle 0|U|1\rangle + \alpha\beta^*\langle 1|U^\dagger|0\rangle). \end{aligned}$$

A diagram that well represents this process makes use of the cap-and-cup diagrams (which recently have been making appearance in quantum computing literature [73, 74] and pop culture [75]) when talking about post-selected processes and can be found in figure 1.9.

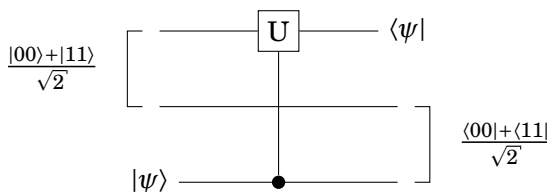


Figure 1.9: Cap-and-cup diagram of the process used to build a self controlled gate utilising post-selection and a standard controlled unitary gate. The bras on the right hand side indicate a successful post-selection upon the state  $\frac{|00\rangle+|11\rangle}{\sqrt{2}}$ .

### 1.3.3 Complexity Theory

The final example in this section will draw upon the work of Scott Aaronson [76]. Suppose one is given a Boolean formula of a large number of variables and we wish to find assignments of the variables that makes the formula true (such problems are commonplace in modern computer science literature under topics such as 3-SAT problems). Provided at least one such setting exists, the problem is easy to solve with post-selection - we set the variables randomly and post-select on the result being true.

Aaronson defined and studied the complexity class  $PostBQP$ , post-selected bounded error quantum polynomial-time, to investigate how powerful the ability to post-select on the outcomes of quantum computers is. Interestingly, in prior work Han, Hemaspaandra, and Thierauf [77] had defined a similar notion for a classical computer, which they defined as  $BPP_{\text{path}}$  and shown that it sits between  $MA$  and  $PP$ , where  $MA$  is a probabilistic generalisation to the famous class  $NP$ , and  $PP$  is the class ‘of problems for which there exists a probabilistic polynomial-time Turing machine that accepts with probability at least 1/2 if and only if the answer is yes.’  $PP$  generalises our notions of finding variable satisfy-ability to be able to answer questions such as do the majority of possible variable inputs satisfy the Boolean formula.

The main result of the Aaronson paper is to show that  $PostBQP = PP$ , the two complexity classes are equivalent. Interestingly, one of the corollaries to this result is that by combining post-selection and quantum mechanics, one has more power (in terms of complexity class structure) than either quantum mechanics alone, or post-selection alone. To phrase another way, post-selection and quantum mechanics together give you the power to not just solve an  $NP$  hard problem, but actually count how many solutions such a problem can have.

One might wonder whether the essential ingredient for quantum computers being so powerful if one is permitted to post-select is the fact that at least some of the underlying dynamics are stochastic (that is, they require a probabilistic description). This has been shown to be unlikely [78], with  $PostBPP$  (the complexity class describing post-selection and classical computation) sitting somewhere between between the first and third levels of the polynomial hierarchy [79]. Since the class  $PP$  sits much higher than this [76, 80], any equivalence would mean the collapse of the polynomial hierarchy - which is considered unlikely but remains an open conjecture.

### 1.3.4 Connections to Causality

Imagine the following set-up. You are outside of a room and your friend, Alice is inside of the room. There is a sliding door that allows you to pass a coin into and out of the room. there is a camera in the room that is recording all events inside the room. Initially, you prepare two coins in the same state, so both heads up or both tails up and then you pass one of the coins into the room. Alice takes the coin from through the sliding door and then tosses it into the air. The camera records the new state of the coin and then she passes the coin back through the sliding door. You then look at the pair of coins. This seems benign enough?

What you witness is that the two coins, who have began in some correlated state, become random relative to one another through the action of Alice flipping them.

Now suppose you repeated this process many times, and show some of the footage to Bob. However, you do not pick which runs of the footage to show him at random, instead you post-select upon footage where the coins were also correlated at the end of the process. This occurs in about one half of the runs, so you aren't short of examples to show to Bob. It is important, by the way, that you as the post-selector have some sort of state dependant filtering of which video tapes to show to Bob. Without this, no paradox will be observed in the subsequent.

What Bob now sees is the coin come into the laboratory in some state, a process that he believes is random occurs to one of the coins and then, magically, the same state that has been issued to Alice is returned by Alice at the end of the experiment.

One had to be serious about the randomising process, but if you are, then one can conclude there has been an influence from future, to the past. Bob sees that the effects of a randomising process are the same as the effect of having not flipped the coins, and one explanation for this is that Alice's output is connected, through an exotic object called a time-like curve, to her input. This paradox, from the point of view of Bob is in essence what forms the close ties between post-selection and simulating exotic causality. We will wait until we properly formalise the two-time formalism in subsequent sections to see this in action.

## 1.4 This Work

This chapter was an introduction and some background on the key topics contained within this thesis. The remainder will be structured as follows. In chapter 2 we will take a dive into a useful formalism for quantum thermodynamics and extend some existing results for unitary evolution to other types of quantum evolution. In chapter 3, we look into causality in quantum mechanics, and prove that coherent control of order cannot be used to violate a causal inequality. In chapter 4, we will be exposed to the two-time formalism in more detail, and show that a local probabilistic notion of causal order does not necessarily enforce a global notion of causality. In chapter 5, we will look at the resource theory of post-selection and find that, while not a quantum resource theory, there is a natural monotone that obeys some altered forms of the desirable properties of a



resource monotone. We will be giving examples and discussions throughout the thesis which I hope will be illustrative and exciting, as well as giving a few open problems and ideas for further research. In chapter 6, we will tie everything we have done so far off with a general discussion and conclusion section. The chapters are based upon papers that are either published (chapters 2 [81], and 4 [64]), To be published in PRL (chapter 3 [82]) or new to online publication (chapter 5). I have also been engaged in other projects during my PhD, and have completed some work ([83]) which will not form any part of this thesis.

## CHANNELS, MEASUREMENTS AND POST-SELECTION IN QUANTUM THERMODYNAMICS

This chapter was completed at the University of Bristol, in conjunction with my supervisor Dr. Anthony J Short. This work is based off the paper ‘Channels, measurements and post-selection in quantum thermodynamics’ [81], which has been published in the journal *Physical Review E*. We have extended and expanded the results of the existing framework of ‘Work extraction and thermodynamics for individual quantum systems’ [19] to quantum channels, quantum measurements and post-selections on quantum systems, and obtained some interesting results concerning the thermodynamics of such objects along the way.

### 2.1 Introduction

What exactly is work? What it good for? And how efficient can it possibly be extracted? These are some of the most pertinent questions that thermodynamics attempts to answer.

Work is taken often in discourse on thermodynamics to mean some form of useful energy [19]. It could be the energy that is generated in a steam engine to move a locomotive between two cities. In a different context, it could be the energy required to lift a weight to a given height in a gravitational field. And it could also be the amount of energy that is spent in a chemical reaction when some chemical bonds become undone and reform themselves into a different arrangement, work is the useful energy in this paradigm spent on the task of molecular synthesis. Whatever the scenario, thermodynamics places bounds on the most amount of work one can extract [84] in that scenario for a state undergoing a physical change.

The vast majority of effort in thermodynamics today from an engineering context goes on under the assumption that the systems and dynamics are classical at the underlying level [85].

One is reminded of the derivation of the laws of ideal gasses by assumptions on such a gas being constructed of many very small bouncing billiard balls [86]. We know of course this to be wrong, but for the vast majority of engineering applications it serves as a good approximation. This work goes against this convention, and instead is one where the underlying rules of the game are quantum mechanical in nature.

Quantum theory, and particularly quantum information theory, has recently received a renewed interest in thermodynamics [87, 12]. These areas include thermodynamics at the nano-scale [88, 89, 90, 91, 92, 93, 94, 19, 95], thermal machines that operate in nano-scale regimes [96, 97, 98, 99, 100, 101], equilibration [102, 103, 104, 105], results pertaining to the relationship between thermodynamic resources [106, 107, 108], Landauer erasure [109, 110, 111], theoretical results concerning the second law of thermodynamics [112, 113, 114, 115, 116] as well as many interesting information theoretic results [117, 118, 119]. One reason for this is an amalgamation of information theory with quantum theory. Information theory was initially constructed, in some ways, to be as abstract from the physical as possible [120]. The content of its character were to treat a message written on paper and one stored in a digital format such as on a hard drive as one-in-the-same. Despite this, answers were found to very earthly and not at all esoteric questions [121]. Consider the strive to answer questions pertaining to communication of such information along a potentially noisy transmitter and receiver set-up, the question of how much space in a disk drive will be needed to store a photograph, and the most efficient way to do so, or perhaps the security of a secret code to a variety of types of hacking. It soon became very apparent to Shannon, and others, that the relevant figures of merit in such problems were entropic in nature [122].

In the remainder of the chapter, we will first see how information theory and its close ties to thermodynamics work in the quantum mechanical formalism in section 2.2, taking into account unitary evolution. Then, in section 2.3 we will consider how things change when quantum channels enter the picture. In section 2.4, we will then use our results for quantum channels to describe the thermodynamics of measurement, in particular deriving a formula for the conditional work extraction. Finally, in section 2.5 we will use our theorem for conditional work extraction to investigate post-selected thermodynamics.

## 2.2 Framework

In this section we will outline the key parts of our formalism, and in doing so our framework for quantum thermodynamics. We will define the various systems and baths in the formalism, make clear what definition we will use for work, heat, entropy and other quantities of interests and, crucially impose the first law of thermodynamics. We consider only quantum systems of finite dimension which we label by  $\rho_S$ , where we will label the working system with a subscript  $S$ . This quantum system is equipped with an arbitrary Hamiltonian  $\hat{H}_S$ . There is also a thermal bath at

temperature  $T$ . The bath is assumed to contain an unlimited supply of states in the thermal state of some bath Hamiltonian,  $\tau_B = \frac{1}{\mathcal{Z}} e^{-\hat{H}_B/T}$ , where  $\mathcal{Z}$  is given by the partition function  $\mathcal{Z} = \text{Tr}(e^{-\hat{H}_B/T})$ . With access to this bath one can generate thermal states no cost, in the nomenclature of resource theories [123] they are ‘free resources’. We will see soon that being ‘free’ also means that such states in the bath *cannot* be used for *useful* extraction of energy from the bath. We define the internal energy of the system appropriately by  $U_s = \text{Tr}[\hat{H}_S \rho_S]$ , the entropy of the system as the Von-Neumann entropy  $S(\rho_S) = -\text{Tr}[\rho_S \ln(\rho_S)]$ , and heat flow as the change in internal energy of the bath  $\Delta Q = \text{Tr}[\hat{H}_B(\tau'_B - \tau_B)]$ , where  $\tau'_B$  refers to the final state of the bath in the transformation. We will use primed systems to mean systems after a thermodynamic operation, and un-primed systems to refer to the states before the transformation. We will also assume that all systems begin uncorrelated, which is a common assumption across many quantum thermodynamic frameworks [12, 124, 125].

### 2.2.1 Implicit Batteries

Often, the notion of work in thermodynamic systems amounts to working implicitly [19], where we only count how much of a quantity is consumed or released by a system through a thermodynamic transformation. This amounts to an quantity defined through the heat flow,

$$(2.1) \quad \Delta Q = \text{Tr}(\hat{H}_B(\tau'_B - \tau_B))$$

and the internal energy change of the system,

$$(2.2) \quad \Delta U = \text{Tr}(\hat{H}_S(\rho'_S - \rho_S))$$

with an imposition of the conservation of energy giving the work done by a process defined as

$$(2.3) \quad \Delta Q + \Delta U + W = 0,$$

to define the work. This identification might not seem useful. It was the case that we wanted work to be identified with some useful energy change going on in the system, but here it appears to be just a numeric quantity that does some book-keeping on behalf of the first law of thermodynamics. Despite this, we can use this framework to do a variety of interesting things. Firstly, we can use this framework to prove a version of the second law of thermodynamics [19], and indeed a third law. We can also show that this framework allows one to associate a quantity called the *free energy*, to the system and the bath which represents the maximum extraditable work from the state of the system [84].

We will leave our justification of the usefulness of the work in this implicit formalism until the next section, and focus here on the mathematical results.

The proof of the second law [19] proceeds as follows. Consider the entropy of the entire system and bath assemblage. Under the assumption that they are initially uncorrelated with one another,

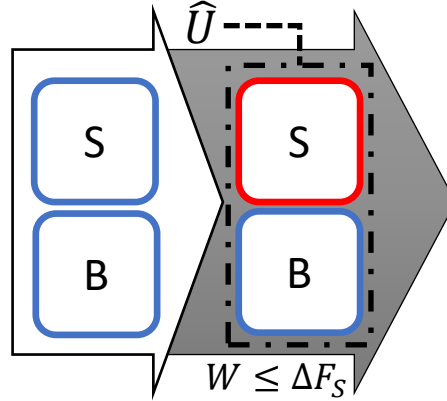


Figure 2.1: Schematic description of a thermal operation in the implicit battery framework. The work is an exogenous quantity which ensures the first law of thermodynamics holds.

we can write

$$(2.4) \quad S(\rho) = S(\rho_S \otimes \tau_B) = S(\rho_S) + S(\tau_B).$$

By employing the sub-additivity of entropy on the primed (after transformation) side

$$(2.5) \quad S(\rho') \leq S(\rho'_S) + S(\tau'_B)$$

and recalling that entropy is a conserved quantity under unitary evolution

$$(2.6) \quad S(\rho_S) + S(\tau_B) = S(\rho) = S(\rho') \leq S(\rho'_S) + S(\tau'_B)$$

we obtain the relation

$$(2.7) \quad \Delta S_S + \Delta S_B \geq 0$$

which is a statement that the entropy of the system and bath are increasing in entropy (because correlations are being built up between them). Hence, in this simple framework, the second law of thermodynamics holds because evolution is unitary.

We now move to investigate how much work we can extract from a given state utilising this framework.

**Definition 2.1.** *The free energy of a system is given relative to a bath at temperature  $T$ , and is given by*

$$(2.8) \quad F(\rho) = U(\rho) - TS(\rho)$$

where  $S(\rho)$  is the Von-Neumann entropy and  $U(\rho) = \text{Tr}[\hat{H}_S \rho]$  is the internal energy of the system. It is of noteworthy remark that the thermal state  $\tau_S = \frac{1}{Z} e^{-\hat{H}_S/T}$  is the unique minimiser of the free energy.

By combining the definition 2.1 and the relation (2.7), we can write

$$(2.9) \quad \Delta F_S + \Delta F_B + W \leq 0.$$

The thermal state of the bath is the minimiser of the free energy, so any change in state of the bath will only ever increase the free energy of this system,  $\Delta F_B \geq 0$ . Applying this to the above inequality we arrive at our statement

$$(2.10) \quad W \leq -\Delta F_S$$

Hence, the maximum work one can extract in this framework is the free energy change of the system undergoing the transformation. Furthermore, any cyclic process which returns the system back to its initial state (so that  $\Delta F_S = 0$ ) can never extract a positive value for work. This is a statement of the second law [126]. Furthermore, we can show in this framework that this bound is achievable, which is to say that there exists an optimal protocol which will extract the free energy difference between final and initial states as work.

First consider the case where  $\rho_{\text{in}}$  and  $\rho_{\text{out}}$  are diagonal in the energy basis and ordered appropriately such that the eigenvalues are in decreasing order (that is, the lower energy eigenstates occur with a higher probability upon a projective energy measurement), and denote them by  $\rho_0$  and  $\rho_{N+1}$  respectively. We next find  $N$  thermal sub-systems  $\rho_1, \dots, \rho_N$  in the bath state which are  $\epsilon$ -close to their predecessor,  $|U(\rho_k) - U(\rho_{k-1})| \leq \epsilon$  for all  $k$ . For the  $k$ 'th step of the protocol, we apply a swap operation between the current state of the system and  $\rho_k$ . After  $N$  such steps we have transformed  $\rho_{\text{in}}$  to  $\rho_{\text{out}}$  as desired, and moreover, since the swap operation is not able to introduce quantum correlations between the system and bath, we have  $\Delta S_S + \Delta S_B = 0$ , and hence  $\Delta F_S + \Delta F_B + W = 0$ . Each sub-system from the bath is initially a minimiser of the free energy, and is transformed to a state which is  $\epsilon$ -close to the original state, so at each step  $\Delta F_B = O(\epsilon^2)$ . So after  $N = O(1/\epsilon)$  steps, we therefore have that  $W = -\Delta F_S - O(\epsilon^2/\epsilon) = -\Delta F_S - O(\epsilon)$ , which can be made arbitrarily close to the bound (2.10) by choosing a small permissible energy difference  $\epsilon$ . If  $\rho_{\text{in}}$  and  $\rho_{\text{out}}$  are not already diagonal in the energy eigenbasis, or are not well ordered as described above, we can apply a unitary transformation at the start and end of the protocol which will render them so, at no extra cost, so will lead to no loss of work.

The third law of thermodynamics can also be generally stated, by noting that in general a finite temperature and zero temperature state have different rank as operators, and that unitary transformations preserve the rank of operators. A more detailed analysis of this setting is found in [127], where a general proof of the third law of thermodynamics is presented, and in [16], where the third law is considered in the context of projective measurements which do change the rank and the costs associated to this are considered.

The main assumptions in these results are that the initial state of the system and bath is uncorrelated (the state can be written in a tensor product form) and that the evolution is unitary. The initially uncorrelated assumption is commonplace in quantum thermodynamics [14,

128], and having it not be true can be shown to posit thermodynamic advantage [124, 125]. It is possible to take advantage of entanglement between the system and the bath in some protocols and extract more work than the second law permits. Other thermodynamic effects beyond work extraction can be shown to be effected by inter-system correlations also. In particular, anomalous heat flows from cold to hot [129], and memory erasure that permits by work extraction instead of heat dissipation when one has lots of information about the system to be erased [117]. These two examples indicate a violation of the Clausius second law formulation, and the Landauer's principle of information erasure [130] respectively.

In this work we want to be fair and imagine ourselves in the situation where we have 'some' system, and want to access tools (the bath) to do something useful (extract work). By assuming the system and bath are initially entangled at some point in their history they would have met (all entangled states are in a sense, non-local [131]) - which is not in the spirit of what we want to do. We could be given the choice between more than one bath to use and we want our results to be independent of this choice (assuming the baths are all at the same temperature). It is indeed true that sometimes things work better together if they have a history of working together, this however is not the physical question we want to answer. We are asking, how well will these things work together - under the proviso they have not done so already in the past.

The latter assumption about unitarity we will tackle later, however for the purposes of the second law we want to avoid cheating by the use of a side-channel (perhaps an energy rich ancilla that we neglect to account for in our bookkeeping). Although it should be stated this is not a shortcoming of the quantum formalism - thermodynamics is a theory which necessarily only works if exogenous sources of energy, particles, entropy are properly accounted for.

A more serious critique of this framework is that it is distinctly nonphysical in the way work is treated. The idea for work is that it is useful energy, and useful is somewhat of a designater without meaning. At the moment, work is not immediately useful, no pistons are being pushed, and no calculation is being done, nothing is getting erased. Simply put, the work here is that which makes the first law of thermodynamics - energy conservation hold. One might wonder if introducing a physical model which utilises this work would add additional restrictions upon the maximal amount of work one can extract. Indeed, outside of this characterisation, it is fair to mention that there is a notion of work that is not necessarily dependable on usefulness, but the reversibility of certain thermal operations [132]. More specifically, the amount of work to be gained from a given thermal operation that takes  $\rho$  to  $\sigma$  depends on how well we can return the state  $\sigma$  to its original form  $\rho$  without investing any additional work (i.e, by making use of a thermal bath only). This framework exhibits an explicit reversal operation in terms of the Petz recovery channel [133]. Results in this framework establishes a quantitative link between the reversibility of thermodynamics processes and a corresponding work gain. It is an interesting open question to investigate how the framework in this chapter and the framework in [132] agree. With the task set to us of providing an operational characterisation to work, we introduce

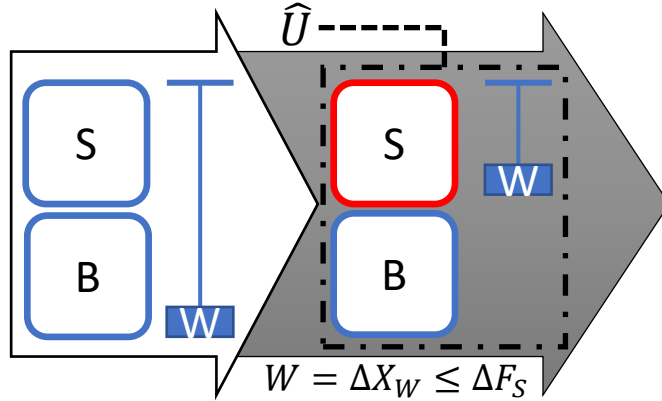


Figure 2.2: Schematic description of the thermal operations in the explicit battery framework. The work is an endogenous property of the system, and is represented by the work done to raise or lower a weight system. This energy can be used at a later time for some other process.

batteries into the picture in a fully quantum way.

### 2.2.2 Explicit Batteries

For our purposes, a work storage device is a battery. A typical battery relies on chemical energy which is released through some breakdown, which is a molecular change [134]. This change can be irreversible, or reversible, or somewhere in between, and carries its own thermodynamic efficiencies which the optimisation of are the focus of large scale research projects costing and producing millions of pounds [135]. In this work, we will use a very simple form of battery, a weight on a string in a gravitational field. The battery is charged by raising the weight and discharged by lowering the weight. This system is sufficiently simple to capture the important ideas present in this work and provides an operational interpretation to the work as being energy of a functional, useful character. To model the weight, we need only one degree of freedom,  $|x\rangle \in \mathcal{H}_W$  (the height), and a Hamiltonian  $\hat{H}_W = mg\hat{x}$ , where we set the product of the mass and the gravitational acceleration to one,  $mg = 1\text{Jm}^{-1}$ . This will allow us to use energies and distances interchangeably.

Now that we have included a battery, we can and it is appropriate to impose a strict form of energy conservation of the combined system. We are going to assume that our system is conserving energy, which means that the dynamics of the system are such that they commute with the Hamiltonian of the system [19]. The allowed unitary transformations therefore satisfy

$$(2.11) \quad [\hat{U}, \hat{H}_S + \hat{H}_B + \hat{H}_W] = 0.$$

Which ensures that energy is not leaving or arriving into our system (other than that provided by the bath and weight). For the same reasons as the implicit battery case, we assume that our system and bath begin uncorrelated, however we will also assume that both the system and the bath are also uncorrelated with the weight. We can therefore write the initial state of the system



as  $\rho = \rho_S \otimes \tau_B \otimes \rho_W$ . We no longer need to define work implicit from heat and internal energy, all three definitions are now provided by the commutation relation (2.11).

$$\begin{aligned}
 \Delta Q &= \text{Tr}[\hat{H}_B \tau'_B - \hat{H}_B \tau_B] \\
 \Delta U &= \text{Tr}[\hat{H}_S \rho'_S - \hat{H}_S \rho_S] \\
 W &= \text{Tr}[\hat{H}_W \rho'_W - \hat{H}_W \rho_W]
 \end{aligned}
 \tag{2.12}$$

from which the first law follows. Our only other assumption for this work is that we will assume that any allowed unitary transformation is transitionally invariant on the weight. To say colloquially, this would mean that a transformation which dropped the weight by 2m has the same effect on the system and gives the same changes in internal energy, heat and work done independent of the initial position of the weight, be it 3m high initially or 1m high. This assumption also implies that the weight has no minimum position. It can rise and fall infinitely. While this may be somewhat nonphysical, work done extending this paradigm [136] has shown that weights that are based on the same idea but using a more physical Hamiltonian such as that of the harmonic oscillator do not cause the fluctuations around the expected heat, internal energy and work done to become too large. Translation invariance on the weight can be surmised by the commutation relation

$$[\hat{U}, \hat{\Gamma}_e] = 0,
 \tag{2.13}$$

where the translation operator  $\Gamma_e$  is given by

$$\hat{\Gamma}_e = e^{-ie\hat{p}_w}
 \tag{2.14}$$

where  $p_w$  is the momentum operator on the space of the weight. To see why this operator is the generator of translations on the weight, it suffices to consider ,

$$\hat{\Gamma}_e |x\rangle = e^{-ie\hat{p}_w} |x\rangle = \int dp e^{-ie\hat{p}_w} e^{-ipx} |p\rangle = \int dp e^{-i(p+e)} |p\rangle = |x+e\rangle
 \tag{2.15}$$

obtained by taking the Fourier transform and inverting. The commutation relation (2.13) prevents us from using the weight system as a thermodynamic resource, for instance as a cold bath - only energy which exists in the system and bath as work goes into raising the weight and only energy which leaves the weight goes into the system and bath and does work.

Other works involving batteries based on two level systems often referred to as a work bit, or *wit* [93], weights with discrete spacing in energy or bounded from below energy spectra [136] exist and are all of interest. The motivation to consider the above model is both in its simplicity, and its comparisons with classical thermodynamics. We are able to visualise the resultant model very easily, and even though it is simple, tweaks to the above idea are sufficiently powerful to consider, for example, multiple conserved quantities [106] and the batteries that store these.

One can also consider time varying Hamiltonian's. The literature on this paradigm is extensive [137, 89, 138], however is not something we consider here. Our results are easy to extend to

the explicit case of the inclusion of a quantum clock to provide a level of control over the thermodynamic cycles [99], and it is worth mentioning that after a thermodynamic process has finished and the weight has been raised if one wishes to disconnect the apparatus and then reconnect it elsewhere to supply the new system with a reservoir of work, a natural method to do this is either to use a time varying Hamiltonian or a coupled quantum control which again, acts just like a clock.

Within this framework we will now also prove that the second law holds, give more details of an optimal work extraction procedure using a weight, and show that the specific state of the weight system is not important when considering unitary interactions. Doing this adds credence to the implicit battery formalism, and will allow us to go beyond either paradigm which are unitary focused to begin to discuss the thermodynamics of channels and measurement. We will begin with a proof of the second law in this framework.

First note that the commutation relation (2.13) implies that in general, any thermodynamic transformation that the system, bath and weight undergo will have the form

$$(2.16) \quad \hat{U} = \int dp \hat{U}_{SB}(p) \otimes |p\rangle \langle p|_W.$$

This means that the system and bath transform,

$$(2.17) \quad \rho'_{SB} = \text{Tr}_W[\hat{U}(\rho_{SB} \otimes \rho_W)\hat{U}^\dagger] = \int dp \mu(p) \hat{U}_{SB}(p) \rho_{SB} \hat{U}_{SB}(p)^\dagger.$$

We have written  $\mu(p) = \text{Tr}_w(\rho_W |p\rangle \langle p|)$ , which is manifestly a probability distribution over momentum. Due to the concavity of the entropy, a probabilistic mixture of unitaries can only ever increase the entropy of the system and bath,  $S(\rho'_{SB}) \geq S(\rho_{SB})$ . As before, combining this with the fact that the initial state is a product state, and utilising the sub-additivity of entropy, we find

$$(2.18) \quad S(\rho_S) + S(\tau_B) = S(\rho_{SB}) \leq S(\rho'_{SB}) \leq S(\rho'_S) + S(\tau'_B)$$

from which, as before it follows that  $W \leq -\Delta F_S$ .

To show that the optimal transformation can be achieved in this framework of quantum thermodynamics in the explicit battery case, we first identify the unitary which achieves the transformation in the implicit battery case.  $\hat{U}$  can be expanded as

$$(2.19) \quad \hat{U} = \sum_{i,j} U_{i,j} |i\rangle \langle j|,$$

with the set  $|i\rangle$  forming an orthonormal basis for the system and bath, in the energy basis with energies  $E_i$ . We then extend this to the weight space by introducing the weight-extended unitary.

**Definition 2.2.** *Given a unitary operator in the implicit battery case (so a unitary acting on the space of system and bath)  $\hat{U}$ , its extension to a weight system with Hamiltonian  $\hat{x}$  is defined as*

$$(2.20) \quad \tilde{U} = \sum_{i,j} U_{i,j} |i\rangle \langle j| \otimes \Gamma_{E_j - E_i}.$$

This method for extending unitaries both preserved the commutation relation (2.13), and gives similar results for the achievable work whenever the initial state has a narrow spread in momentum about the zero of momentum. Asserting this will mean that we can use an identical protocol to the one outlined in the implicit case to show optimality of the protocol.

We first begin by writing  $\tilde{U}$  in the momentum basis as,

$$(2.21) \quad \tilde{U} = \int dp \sum_{i,j} U_{i,j} e^{-i(E_j - E_i)p} |j\rangle \langle i| \otimes |p\rangle \langle p|.$$

The state of the system and bath after applying  $\tilde{U}$  is given by

$$(2.22) \quad \rho'_{SB} = \int dp \mu(p) U(p) \rho_{SB} U(p)^\dagger,$$

where

$$(2.23) \quad U(p) = \sum_{i,j} U_{i,j} e^{-i(E_j - E_i)p} |j\rangle \langle i|.$$

Consider as a first case that the initial state of the system and bath is quasi-classical, which is to say that  $\rho_{SB} = \sum_n p_n |n\rangle \langle n|$ , and assume that the unitary is a permutation among energy levels,  $U_{i,j} = \delta_{i,\pi(j)}$  with  $\pi$  a permutation matrix. It is not hard to see that in this case,  $U(p) \rho_{SB} U(p)^\dagger = U \rho_{SB} U^\dagger$  for all momentum, and hence the work-optimally relation is exact. If instead, we have a general state of system and bath, a general unitary and a weight state that is tightly peaked in momenta, we can show that the work-optimally is within an error that can be made arbitrarily small. Suppose that

$$(2.24) \quad \int_{-c}^c dp \mu(p) \geq 1 - \delta$$

for some small  $\epsilon, \delta$ . The induced transformation can be calculated from (2.17), and the trace-distance between this state and the desired final state from the implicit battery case,  $\hat{U} \rho \hat{U}^\dagger$  can then be bounded as follows

$$(2.25) \quad \begin{aligned} D(\rho'_{SB}, \hat{U} \rho_{SB} \hat{U}^\dagger) &= \frac{1}{2} \|\rho'_{SB} - \hat{U} \rho_{SB} \hat{U}^\dagger\|_1 \\ &\leq \frac{1}{2} \int dp \mu(p) \|U(p) \rho_{SB} U(p)^\dagger - \hat{U} \rho_{SB} \hat{U}^\dagger\|_1 \\ &\leq \delta + \frac{1}{2} \int_{-c}^c dp \mu(p) \|U(p) \rho_{SB} U(p)^\dagger - \hat{U} \rho_{SB} \hat{U}^\dagger\|_1 \\ &\leq \delta + \frac{1}{2} \int_{-c}^c dp \mu(p) (4|p| \|\hat{H}_{SB}\| + O(p^2)) \\ &\leq \delta + O(\epsilon) \end{aligned}$$

where, for the third step we have expanded a power series for the unitary  $U(p)$ ,

$$(2.26) \quad U(p) = U - ip[U, \hat{H}_{sb}] + O(p^2)$$

and we have used the fact that for any density operator  $\|H\rho\|_1 \leq \|H\|$ . Hence, by using a weight state with sufficiently small  $\epsilon$  and  $\delta$ , we can make the final state of the system and bath arbitrarily close in energy to the desired final state. As the system and bath are defined in such a way to have bounded spectrum, this means that their final expected energy is arbitrarily close to the desired value in the implicit battery case. By total energy conservation (2.11), the amount the weight is lowered or raised, the work, is also arbitrarily close to the value when the battery is implicit.

As a concrete example, consider the state of the weight to be a pure ‘top-hat’ function of width  $L$ .

$$(2.27) \quad |\phi_L\rangle_W = \frac{1}{\sqrt{2L}} \int_{-L}^L dx |x\rangle.$$

one can show through the Fourier transform that

$$(2.28) \quad \mu(p) = \frac{1}{\pi L} \frac{\sin^2(pL)}{p^2}$$

the ‘sinc-squared’ function. By taking  $\epsilon = \frac{c}{\sqrt{L}}$  for some constant  $c$ , we find that

$$(2.29) \quad \begin{aligned} \int_{\epsilon}^c dp \mu(p) &\geq 1 - 2 \int_0^{\infty} dp \mu(p) \\ &\geq 1 - \frac{2}{\pi L} \int_{\frac{c}{\sqrt{L}}}^{\infty} dp \frac{1}{p^2} \\ &\geq 1 - \frac{2}{\pi c \sqrt{L}} \end{aligned}$$

and hence both  $\epsilon \rightarrow 0$  and  $\delta \rightarrow 0$  as  $L \rightarrow \infty$ . It is simple to construct thermodynamic protocols, of which we shall now sketch out the details of the most important protocol, the Carnot cycle. For this, we introduce two baths, one held at a temperature  $T = T_h$  which will be the hot bath and one at  $T = T_c$ , the cold bath. We also have our working quantum system and weight system. Imagine that the system is initially in thermal equilibrium with the cold bath, so that its internal energy and entropy can be written as  $U_c$  and  $S_c$  respectively. One stroke of the cycle is given as follows

1. First, take the state and decouple from the cold bath. Using the above protocol, attach it to the hot bath. In the best case, we can extract work  $W_1 = -\Delta F_S = (U_c - T_h S_c) - (U_h - T_h S_h)$ .
2. Second, decouple the state from the hot bath and attach it to the cold bath. This extracts work  $W_2 = (U_h - T_c S_h) - (U_c - T_c S_c)$ .

applying the first law to the work

$$(2.30) \quad W = W_1 + W_2 = (T_h - T_c)(S_h - S_c),$$

we find that in the best case,

$$(2.31) \quad U_h - U_c = Q_h - W,$$

with an efficiency given by the work out per stroke, divided by the wasted energy, the heat flow

$$(2.32) \quad \frac{W}{Q_h} = 1 - \frac{T_c}{T_h}.$$

which is the famous Carnot result, holding well over 200 years after its initial derivation within the scope of classical thermodynamics.

The fact that the results for the implicit and explicit case of batteries are identical (at least, for the case of unitary evolution) offers enhanced justification for utilising the implicit case in practical problems. However, often during a quantum process, and in particular in many thermodynamic processes where the quantum system is open and evolution is not unitary. We shall see that insights from the explicit battery case also prove useful in understanding the thermodynamics of channels, measurements and post-selected systems.

## 2.3 Channels

We will now move on from our exclusive treatment of unitary quantum mechanics, into the realms of channels [139, 140]. Recall that a quantum channel as defined in (1.6) is a completely positive trace preserving map from density matrices to density matrices. As well as being of interest in its own right, this section of the thesis is important for understanding the subsequent sections concerning measurements and eventually post-selection.

In this section, we will consider what thermodynamic advantage (if any) quantum channels posit over what is available in the unitary framework above, and define properly the thermodynamics costs of such objects. Consider you are granted a single use of a quantum channel  $C$  in a thermodynamic cycle. This channel is applied to a target system  $T$ , in  $\mathcal{H}_T$ , with Hamiltonian  $H_T$ . we will also allow you access to an ancilla  $A$ , in  $\mathcal{H}_A$ . The working quantum system from before is then just the tensor product of  $T$  and  $A$ . We note that the system  $A$  is not the system that acts to dilate the channel to a unitary realisation, it is some other ancilla that exists in the system. In other words, it is an endogenous system (the system that dilates the channel on the other hand will be considered exogenous). It will be helpful to define the action of the channel on the system, leaving the ancilla untouched by  $C_S = (\mathbb{1} \otimes C)$ .

Such scenarios arrive in the study of open quantum systems [141]. We could realise the channel, either through some quantum computation algorithm [142], or a physical scenario - consider molecules which can be intervened upon by an experimenter. The molecules, provided they are well isolated from their environment would be undergoing unitary evolution, however should an experimenter choose to intervene in the interaction, say by shining some light upon the molecules, then the effect of the external system (the light) upon the molecules would be described by a quantum channel.

Quantum channels are also being considered to provide primitive objects for quantum computation. In [143], the authors consider a measurement based approach to quantum computation, which uses measurements instead of loss to simulate a variety of different quantum channels on photons.

As their name suggests, quantum channels are analogous objects to the classical counterparts in communication theory - coming with quantum counterpart versions of famous classical structures such as the channel capacity [144] and famous theorems such as the source coding theorem [145, 146]. As the quantum technologies industry grows, there is hope that certain communication scenarios, particularly ones where privacy is important will utilise quantum channels as their communication mode. Exciting progress includes city-wide [147], and more recently space-scale [148] realisations of communication through quantum channels.

Returning to thermodynamics, We will consider the implicit battery framework first, in which the channel is treated as a primitive operation. We will then show that the same results are obtained in an explicit framework in which the channel is implemented via a unitary on ancillary system, through strictly energy-conserving interactions with an explicit battery.

We will quantify the work benefit of the channel by the maximum work that can be extracted in a cyclic process, in which the channel is applied only once. Cyclic here, means that all endogenous variables in the system must be returned to the same state at the end of the protocol as they are at the beginning. A well studied generalisation [149, 150] to this framework includes approximately cyclic processes, where the states can be returned at the end of the protocol to within some error, however we will only consider exactly cyclic processes. We allow the entirety of the previous unitary framework, with a single thermal bath at temperature  $T$ . There are two natural frameworks in which to proceed. The first, which the interested reader can find more about in [81] is to consider that the channel is free to perform (perhaps, the gadget which implements the channel comes with its own power supply). An obvious downside to this framework is that even unitary channels would give a thermodynamic advantage, so in keeping with the framework of the first section, we imagine the channel shares the same power supply as the unitary operations - namely, the weight system.

A general procedure for work extraction in this framework is the following:

1. The system and bath are initially uncorrelated,  $\rho = \rho_S \otimes \tau_B$ .
2. The channel is applied to the target system, giving  $\sigma = C_S[\rho_S \otimes \tau_B]$ .
3. A unitary interaction resets the state to  $\sigma' = \hat{U}\sigma\hat{U}^\dagger$ , where  $\sigma'_S = \rho_S$ .

specifying a protocol  $P$  includes picking the desired target, the desired thermal bath, an ancilla and as well the unitary  $\hat{U}$ . The total work is given by

$$(2.33) \quad W_{C,P}^{\text{total}} = W_1 + W_2 + W_3$$

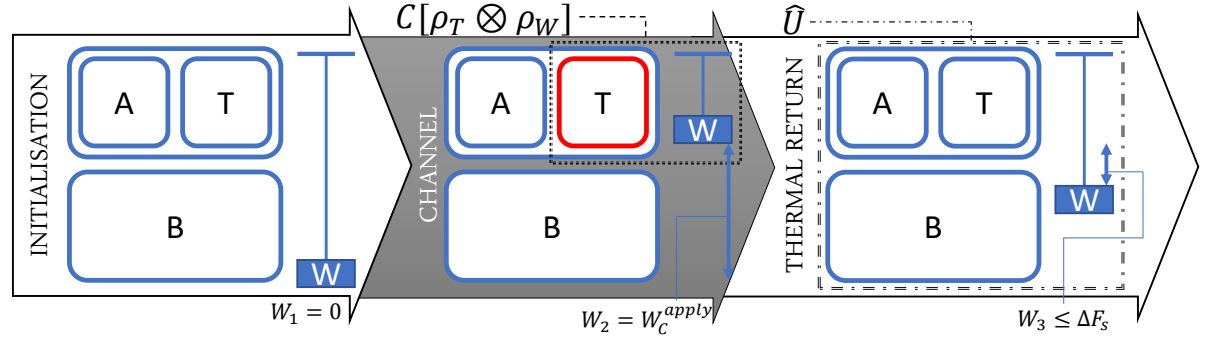


Figure 2.3: Schematic description of the protocol  $P$ , which consists of a channel application and a unitary reset. A single use of a quantum channel is used to provide a work advantage in a cyclic protocol. As in the main text, the system is composed of an ancilla and target  $S = A \otimes T$ , the bath is represented by  $B$  and the weight system by  $W$ .

where the subscript is used to index the stages of the protocol which is occurring, e.g.  $W_1$  is the work associated to picking  $\rho_S, \tau_B$  uncorrelated,  $W_2$  is the work associated to the channel application and  $W_3$  is the work associated to the unitary resetting.

**Definition 2.3.** *The work gained when applying the channel  $C$  to the target system in state  $\rho_T$ , part of a larger system in state  $\rho_S$ , is given by*

$$(2.34) \quad \begin{aligned} W_C^{apply} &= -\Delta U_{\rho_T \rightarrow C[\rho_T]} = \text{Tr}[\hat{H}_T \rho_T - \hat{H}_T C[\rho_T]] \\ &= \text{Tr}[\hat{H}_S \rho_S - \hat{H}_S C_S[\rho_S]] \end{aligned}$$

By making use of definition (2.3) and the results concerning the optimal unitary protocol, the work in each step can be bounded,

$$(2.35) \quad \begin{aligned} W_1 &= 0 \\ W_2 &= W_C^{apply} = U(\rho_S) - U(\sigma_S) \\ W_3 &\leq F(\sigma_S) - F(\rho_S) \end{aligned}$$

Hence, from (2.10):

$$(2.36) \quad \begin{aligned} W_{C,P}^{\text{total}} &\leq T(S(\rho_S) - S(\sigma_S)) \\ &= T(S(\rho_S) - S(C_S[\rho_S])) \\ &\leq T(S(\rho_T) - S(C[\rho_T])) \end{aligned}$$

where the last step follows from the monotonicity of the mutual information under the action of (local) channels. We note here that for any  $\rho_T$  we can consider that no ancilla actually gets used by the protocol  $P$  and so  $\rho_T = \rho_S$ , and that  $U$  is a thermodynamically efficient protocol for transforming  $C_S[\rho_T]$ . Hence, there exist protocols which come as close as one would like to saturating the bound (2.36).

We can optimise over the protocols  $P$ , to find the benefit of the channel.

**Theorem 2.1.** *The work benefit of a quantum channel  $C$  is given by*

$$(2.37) \quad W_C^{\text{total}} = \sup_P W_{C,P}^{\text{total}} = \max_{\rho_T} T(S(\rho_T) - S(C[\rho_T]))$$

Note that  $0 \leq W_C^{\text{total}}$  for all channels. One can see this since

$$(2.38) \quad \max_{\rho_T} T(S(\rho_T) - S(C[\rho_T])) \geq T(S(\mathbb{1}/d) - S(C[\mathbb{1}/d])) = T(\ln(d) - S(C[\mathbb{1}/d]))$$

In other words, the fact that the maximally mixed state already has maximum entropy means that the  $W_C^{\text{total}}$  is a zero or positive quantity. However, we will only call a quantum channel useful from a thermodynamic perspective if  $0 < W_C^{\text{total}}$ , which from the theorem above means that it has the ability to reduce the entropy of some state by its action. Before we attempt to justify this definition with an explicit model, it will be interesting to identify some classes of channels that have interesting relations to the total work.

**Corollary 2.1.** *A channel provides no work benefit if and only if it is unital.*

A unital quantum channel  $C_{\text{un}}$  are those whose action on the maximally mixed state is identity,  $C_{\text{un}}[\mathbb{1}/d] = \mathbb{1}/d$  [151]. These include unitary channels, or mixtures of unitary channels, but there are also other possibilities not in either of those classes [152]. In two dimensions these classes coincide [153] but in higher dimension, this is no longer the case. We first prove that the unital channels cannot decrease the entropy of the state, by considering first the relative entropy

$$(2.39) \quad D(\rho||\sigma) = \text{Tr}(\rho \ln(\rho) - \rho \ln(\sigma)).$$

It can be shown that for any quantum channel, the relative entropy carries the monotonicity property,  $D(\rho||\sigma) \geq D(C[\rho]||C[\sigma])$ . By setting  $\sigma = \mathbb{1}/d$ , for unital channels we find  $D(C_{\text{un}}[\rho]||\mathbb{1}/d) \geq D(\rho||\mathbb{1}/d)$ . Expanding out leads to

$$(2.40) \quad \begin{aligned} D(\rho||\mathbb{1}/d) &\geq D(C_{\text{un}}[\rho]||\mathbb{1}/d) \\ \text{Tr}(\rho \ln(\rho) - \rho \ln(\mathbb{1}/d)) &\geq \text{Tr}(C_{\text{un}}[\rho] \ln(C_{\text{un}}[\rho]) - C_{\text{un}}[\rho] \ln(\mathbb{1}/d)) \\ \text{Tr}(\rho \ln(\rho)) + \text{Tr}(\rho) \ln(d) &\geq \text{Tr}(C_{\text{un}}[\rho] \ln(C_{\text{un}}[\rho])) + \text{Tr}(C_{\text{un}}[\rho]) \ln(d) \\ \text{Tr}(\rho \ln(\rho)) &\geq \text{Tr}(C_{\text{un}}[\rho] \ln(C_{\text{un}}[\rho])) + \\ S(C_{\text{un}}[\rho]) &\geq S(\rho) \end{aligned}$$

and hence  $W_{C_{\text{un}}}^{\text{total}} = 0$ . For the only if part, note that a non unital channel must satisfy  $C[\mathbb{1}/d] \neq \mathbb{1}/d$ . As the maximally mixed state is the unique state with maximum entropy for a given dimension, it must therefore be the case that for some state (namely, the maximally mixed state) that  $S(C[\rho]) \leq S(\mathbb{1}/d)$  and hence the work total is positive.

Another particularly interesting class of channels in this context are catalytic channels [154].



**Definition 2.4.** A channel  $C$  is catalytic if there exists a unitary  $\hat{V}$  and a state  $\rho_Z$  such that

$$(2.41) \quad \begin{aligned} C[\rho_T] &= \text{Tr}_Z(\hat{V} \rho_T \otimes \rho_Z \hat{V}^\dagger) \\ \rho_Z &= \text{Tr}_T(\hat{V} \rho_T \otimes \rho_Z \hat{V}^\dagger) \end{aligned}$$

so that the ‘catalyst’ system is returned as it was before the channel was applied.

Note that our notions of catalysis are exact. There exist paradigms [154, 155, 23] which take into account inexact catalysis where the catalyst is returned up to an additive error  $\epsilon$ . As pointed out in [156], how one calculates the error becomes important if one includes any error. A natural and operationally motivated error could be the trace distance between the catalyst before and after a transformation. However, surprisingly states that are close in trace distance may have quite different thermodynamic properties, and posit different thermodynamic advantages. One can envisage a scenario where the catalyst has returned to a state that is close in trace distance to the initial state, but lifts limitations posited by the second law by acting as an entropy sink or work source. This leads to what has been called embezzlement of work [154] and is something we wish to avoid in this work. So as before for our notion of a cyclic protocol, and with quantifying work in mind, we want to be fair across the board and so demand exact catalysis for the following study.

Given that a catalytic implementation of a channel does not need the ancilla qubit to be reset after the operation, We will show in the following section that this means that the total work extracted by the channel is not positive (there is no work benefit, it cannot raise the weight system into a higher energy). From corollary (2.1), we can thus deduce that catalytic channels are unital.

This raises the interesting question of whether all unital channels are catalytic in this sense. We will now show that this is not the case. We will show this through a direct counter example - and the example we will pick is the Werner-Holevo channel [157] for a three dimensional system (on the other had, it is known that for  $d = 2$  all unital channels are mixtures of unitaries, and hence unital [158]). The Werner-Holevo channel can be expressed as

$$(2.42) \quad \rho \rightarrow C_{\text{wh}}[\rho] = \frac{1}{2}(\text{Tr}[\rho]\mathbb{1} - \rho^T)$$

where  $\rho^T$  represents the transpose of  $\rho$  in a basis. This channel can indeed be verified to be unital. It is also not a unitary channel, consider applying the channel to a projector  $\pi = |j\rangle\langle j|$ , when the transpose is taken in the  $|j\rangle$  basis. Note that the eigenvalues of the  $\pi$ ,  $C_{\text{wh}}[\pi]$  differ. It is an essential fact for the following proof that  $C_{\text{wh}}[\rho]$  is extremal in the space of channels, and therefore cannot be expressed as a mixture of different channels [159].

Suppose that there was a catalytic implementation of  $C_{\text{wh}}$ ,

$$(2.43) \quad \begin{aligned} C_{\text{wh}}[\rho_T] &= \text{Tr}_Z[\hat{V} \rho_T \otimes \rho_Z \hat{V}^\dagger] \\ \rho_Z &= \text{Tr}_T[\hat{V} \rho_T \otimes \rho_Z \hat{V}^\dagger] \end{aligned}$$

for some unitary  $\hat{V}$ . Since  $\rho_Z$  is unchanged by a catalytic  $C_{\text{wh}}$ , the unitary  $\hat{V}$  cannot take states from within the support of  $\rho_Z$  outside of the support. Hence, WLOG we can restrict  $Z$  to the support of  $\rho_Z$ . For any state  $|\psi\rangle_Z$ , then there exists an  $\epsilon$  such that  $\rho_Z = \epsilon|\psi\rangle\langle\psi| + (1-\epsilon)\sigma_Z$ , where  $\sigma_Z$  is a valid state. This procedure of picking the closest pure state to a density matrix has been explored well in the literature [160]. This construction implies

$$(2.44) \quad C_{\text{wh}}[\rho_T] = \epsilon C_\psi[\rho_T] + (1-\epsilon)C_{\sigma_Z}[\rho_T]$$

where

$$(2.45) \quad \begin{aligned} C_\psi[\rho_T] &= \text{Tr}_Z[\hat{V}\rho_T \otimes |\psi\rangle\langle\psi| \hat{V}^\dagger] \\ C_{\sigma_Z}[\rho_T] &= \text{Tr}_Z[\hat{V}\rho_T \otimes \sigma_Z \hat{V}^\dagger] \end{aligned}$$

are both valid channels. As  $C_{\text{wh}}$  is extremal, it is the case that  $C_{\text{wh}_\psi}[\rho_T] = C_{\text{wh}}[\rho_T]$  for all  $|\psi\rangle$ . This kind of independence on the ancilla further implies that  $\hat{V}$  is a product unitary  $\hat{V}_T \otimes \hat{V}_Z$  [161]. This is a contradiction with the fact that  $C_{\text{wh}}$  is not unitary. Therefore,  $C_{\text{wh}}$  is not a catalytic channel. This is best surmised by the following :

**Corollary 2.2.** *All catalytic channels are unital. However, there exist unital channels which are not catalytic.*

Overall, we have shown the following interesting triad of implications between quantum channels that are not useful for work generation, channels that are unital and channels that have a catalytic implementation

$$(2.46) \quad \text{Catalytic} \implies \text{No work benefit} \iff \text{Unital}.$$

Given that many channels are not unital, and so provide a work benefit with heat being transferred into work as a system undergoes a cyclic process, we must explore constancy with the second law of thermodynamics. If this does not hold, the framework is wrong. Constructing an explicit model of the channel, via a unitary interaction between the target system and an additional ancilla which we label  $Z$  to distinguish from the ancilla  $A$  which forms part of the working system, as in figure 2.3. The channel then can be written as

$$(2.47) \quad C[\rho_T] = \text{Tr}[\hat{V}\rho_T \otimes \rho_Z \hat{V}^\dagger]$$

where  $\rho_Z$  is the state of the additional ancilla and  $\hat{V}$  is unitary. For simplicity, we will assume  $\hat{H}_Z$  is trivial. Doing this ensures we only need to be concerned with the entropies, energies can always be added in later if we wish. Due to the sub-additivity of entropy, we find

$$(2.48) \quad \Delta S_T + \Delta S_Z \geq 0$$

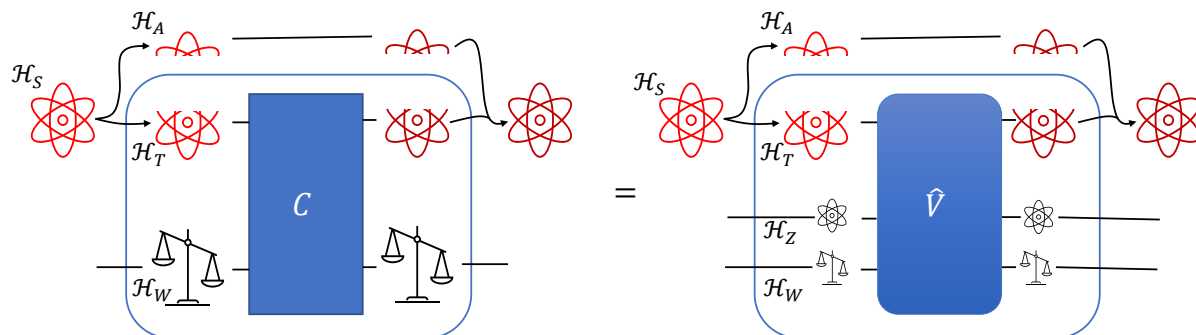


Figure 2.4: Illustration of the dilation theorem, where a quantum channel is realised by a unitary operation on a larger system.

and so by following the same relative entropy argument as above we find that  $W_{C,P}^{\text{total}} \leq -T\Delta S_Z$ . If we want to reset this ancilla  $Z$  to its initial state to make a truly cyclic protocol over all systems, then by employing the unitary framework and the use of a thermal bath we find

$$(2.49) \quad W_{C,P}^{\text{reset } Z} \leq -T\Delta S_Z$$

where  $\Delta S_Z$  is the same entropy change that the ancilla underwent when the channel was initially applied. It then follows that

$$(2.50) \quad W_{C,P}^{\text{total}} + W_{C,P}^{\text{reset } Z} \leq 0$$

in accordance with the second law.

We do not consider the thermodynamic costs of setting up the initial state of the system  $\rho_S$ , or the device that implements the channel  $\rho_Z$ , before the protocol  $P$  begins. Our justification for this is that in the former case, the protocol is cyclic, and can in principle be repeated many times with identical copies of the channel. It is natural to assume that the thermodynamic cost of an initialisation operation therefore becomes more and more negligible as the protocol runs many times. In the latter case, we have considered the cost of resetting the device so the same argument applies.

A subtle issue is that if either of these two states is not full rank. If not, it may be impossible to reconstruct them exactly via the unitary framework. However, for the system considering a series of full rank states that come arbitrarily close to the optimum state, we can see the optimum in (2.1) becomes the maximum over all states. We can also consider a series of devices in much the same manner, and achieve optimality.

We now proceed with seeking justification for the implicit battery assertions with an explicit model, for the case of channels.

As before, the quantum channel is achieved by a unitary evolution  $\hat{V}$  on the target and ancilla system. We once again assume that the ancilla system is degenerate in energy. Expanding  $\hat{V}$  in

the energy basis gives

$$(2.51) \quad \hat{V} = \sum_{a,b,c,d} \alpha_{a,b}^{c,d} |a\rangle \langle b| \otimes |c\rangle \langle d|$$

where  $\alpha_{a,b}^{c,d}$  are complex numbers describing the state transition amplitudes within the Hilbert space  $\mathcal{H}_T \otimes \mathcal{H}_Z$ . As in the implicit battery case, we construct an energy conserving unitary  $\tilde{V}$  with a simple construction. We imagine once again that this is achieved by lowering or raising a weight system.

**Definition 2.5.** *The extension of a unitary  $\hat{V}$  used in the implicit case to implement the quantum channel in the explicit case is given by*

$$(2.52) \quad \tilde{V} = \sum_{a,b,c,d} \alpha_{a,b}^{c,d} |a\rangle \langle b| \otimes |c\rangle \langle d| \otimes \Gamma_{E_b - E_a}$$

where  $\Gamma$  is the translation operator for the weight system, defined in (2.14).

Clearly  $\tilde{V}$  is unitary since (by letting  $i = (a, c)$  and  $j = (b, d)$  and abusing notation slightly) we have that

$$(2.53) \quad \tilde{V}^\dagger \tilde{V} = \sum_{i,j} \alpha_{i,j}^* |i\rangle \langle j| \otimes \Gamma_{E_i - E_j}^\dagger \sum_{k,l} \alpha_{k,l} |k\rangle \langle l| \otimes \Gamma_{E_l - E_k}$$

$$(2.54) \quad = \sum_{i,j,l} \alpha_{i,j}^* \alpha_{i,l} |j\rangle \langle l| \otimes \Gamma_{E_l - E_j}$$

$$(2.55) \quad = \sum_{i,j} \delta_{j,l} |j\rangle \langle l| \otimes \Gamma_{E_l - E_j} = \sum_i |i\rangle \langle i| \otimes \Gamma_0 = \mathbb{1}.$$

It is also simple to check that  $\tilde{V}$  commutes with the total Hamiltonian of the system, ancilla and weight;

$$(2.56) \quad [\tilde{V}, \hat{H}_{SBW}] = (\hat{H}_S B + \hat{H}_W) \tilde{V} - \tilde{V} (\hat{H}_S B + \hat{H}_W) = \sum_{i,j} \alpha_{i,j} E_i |i\rangle \langle j| \otimes \Gamma_{E_j - E_i} + \alpha_{i,j} |i\rangle \langle j| \otimes \hat{H}_W \Gamma_{E_j - E_i} - \sum_{i,j} \alpha_{i,j} E_j |i\rangle \langle j| \otimes \Gamma_{E_j - E_i} + \alpha_{i,j} |i\rangle \langle j| \otimes \Gamma_{E_j - E_i} \hat{H}_W$$

which can be seen to be equal to zero since we have that  $\Gamma_a \hat{H}_W \Gamma_a^\dagger = \hat{H} - a \mathbb{1}$  which implies that  $\Gamma_a \hat{H}_W = \hat{H} \Gamma_a - a \mathbb{1} \Gamma_a$ , making the commutation relation  $[\tilde{V}, \hat{H}_{SBW}] = 0$  hold by setting  $a = E_j - E_i$ . Also, clearly  $\tilde{V}$  has the weight translation invariance property, since the gamma operator (The only part of the operator acting on  $\mathcal{H}_W$ ) itself does.

As the device interacts with the weight system, it's action on the target system will in general depend upon the state of the weight. Given a weight in state  $\rho_W$  the action of the channel is

$$(2.57) \quad C_{\rho_W}[\rho_T] = \text{Tr}(\tilde{V}(\rho_T \otimes \rho_Z \otimes \rho_W) \tilde{V}^\dagger).$$

By considering the device systems  $T$  and  $Z$  a straightforward application of the proof from the unitary case carries over, so that a weight state with a narrow spread in momentum about

$p = 0$  will implement the channel  $\epsilon$  close to the channel from the implicit battery case. Note that such initial states of the weight are also those required to perform arbitrary transformations of a system and bath with optimal thermodynamic efficiency. As an example, consider the state of the weight

$$(2.58) \quad |\psi_L\rangle_W = \frac{1}{\sqrt{2L}} \int_{-L}^L dx |x\rangle$$

with  $\rho_W = |\psi_L\rangle\langle\psi_L|$ . It then follows that  $\lim_{L \rightarrow \infty} C_{\rho_W}[\rho_T] = C[\rho_T]$  exactly. Then, in the infinite  $L$  limit, the change in average energy will be similar to  $\Delta U_{\rho_T \rightarrow C[\rho_T]}$ . As energy is conserved overall, the shift in average energy of the weight will therefore be given by definition 2.3 as desired. As in the unitary case, we can thus assign the implicit and explicit battery conditions consistently.

## 2.4 Measurements

As we have seen, the absence of an operational interpretation for work in the implicit model of unitary and channel thermodynamics *can* be grounded in an operational model; the explicit paradigm, constantly. What happens when we consider measurements?

Often works in this direction either do not account for the process of measurement, or have difficulty assigning meaningful thermodynamic quantities to it. While the deterministic evolution of the Schrödinger equation can and is considered with relative ease in thermodynamics, measurement, being a probabilistic process, requires a more careful treatment - it is statistical phenomena on top of statistical phenomena! It is not immediately clear how to impose the first, second, or third law of thermodynamics at the measurement step. The concept of metroropy [162] has been introduced for calculating work extraction due to measurement. Metroropy is given by  $M = E - \min E'$ , where  $E$  and  $E'$  are the pre and post-measurement energy of the quantum state. A disadvantage of the metroropy approach is that the state in a thermal cycle can be an unknown, and to calculate in this case additional measurements must be performed before and after the measurement step - and so it is unclear what quantum mechanical effects are allowed to enter the picture, rendering metroropy an energy associated with a stochastic state transition. The explicit and operational picture provided by the weight system can alleviate some of these difficulties. In the explicit framework both quantum coherence and knowledge of measurement outcomes can be included simultaneously.

In fact, most of the state of the art quantum measurements of energy are taken in the two-point measurement scheme [163, 164]. The procedure can be outlined as follows.

Consider an isolated quantum system initially prepared in some equilibrium state. The system is subjected to a projective energy measurement at times  $t = 0$  and  $t = \tau$ . In the intermediate times, an external force acts on the system and changes its energy. Let this force be totally characterised by the parameter  $\lambda_t$ . The energy measurements then return the values  $E_n^{\lambda_0}$  for the first and  $E_m^{\lambda_\tau}$  for the second measurements, where  $n$  and  $m$  are used to index the respective

energy levels of the initial and final Hamiltonian. Thermal and quantum randomness render the measured energy difference  $E_m^{\lambda_\tau} - E_n^{\lambda_0}$ , which can be interpreted as the work done on the system through the protocol as a stochastic variable. Unfortunately, such strategy has a considerable drawback from the quantum side, in that, by performing pairs of strong projective measurement, all quantum coherence in the energy eigenbasis are removed, *de facto* removing the possibility of quantum interference to take place [165].

Another option then is to use a two-time measurement procedure, a small alteration to the aforementioned strategy. Instead of preparing our system in an equilibrium state, we instead prepare a probe which is centred about the zero in its position co-ordinate. At time  $t = 0$ , the probe and the system are coupled by the interaction Hamiltonian

$$(2.59) \quad H_{int} = \delta(t - 0)p \otimes H_s.$$

This coupling will move the state of the probe by a position  $E_n^{\lambda_0}$ . Then the same forcing strategy is applied to the system, which changes its energy, after which a second interaction Hamiltonian is applied

$$(2.60) \quad H_{int} = -\delta(t - \tau)p \otimes H_s$$

which moves the probe system by an amount  $-E_m^{\lambda_\tau}$ . Finally the position of the probe is read, by a projective measurement on the position basis, which reveals only the difference (in contrast to before which reads the absolute value of each component)  $E_m^{\lambda_\tau} - E_n^{\lambda_0}$ . Identifying work with this difference leads to work being distributed with probability

$$(2.61) \quad p(w) = \sum_{n,m} p_n p_{n,m} \delta(w - (E_m^{\lambda_\tau} - E_n^{\lambda_0}))$$

where  $p_n$  is the probability to obtain energy  $E_n^{\lambda_0}$ ,  $\langle \phi_n | \rho_{i,n} | \phi_n \rangle$  with  $|\phi_n\rangle$  an energy eigenstate of the Hamiltonian at time  $t = 0$ . Also,  $p_{n,m} = |\langle \phi_n | U_{\tau,0} | \tilde{\phi}_m \rangle|^2$  are the jumping or hopping probabilities for the system where here  $|\tilde{\phi}_m\rangle$  are the energy eigenstates of the final system Hamiltonian. From this procedure, the famous quantum fluctuation relation first appearing in [166] and expanded on in [167] can be derived, by once again providing a thermal equilibrium state as the input. Provided the probe is not read in the intermediate times, this will permit genuine quantum interference, however if the experimenters decide to now measure the system in the energy eigenbasis by means of a projective energy measurement, the value  $E_m^{\lambda_\tau}$  can be learned and since we know this value and its difference from the initial energy, we are also able to obtain  $E_n^{\lambda_0}$  - an energetic version of the quantum eraser.

We now have two options, we can randomise the system before a second projective energy measurement can take place and attempt to mask the information about  $E_m^{\lambda_\tau}$ , or do something different entirely [66] which hopes to reveal the quantum mechanical properties of work. We outline a procedure for the second case briefly here. If instead of preparing a single probe and interacting with the system twice, we prepare two probes and interact them with the system

once each, we can exploit the properties of quantum entanglement to ensure we can measure the quantum mechanical effect in a thermodynamics state transition. We prepare two probes in the entangled state  $|q_1 - q_2 = 0, p_1 + p_2 = 0\rangle$ , where  $q_1, q_2$  are position coordinates of the probes and  $p_1, p_2$  their momenta. By applying the interaction Hamiltonian

$$(2.62) \quad H_{int} = \delta(t - 0)p_1 \otimes H_s - \delta(t - \tau)p_2 \otimes H_s$$

we notice that the value of  $E_n^{\lambda_0}$  and  $E_m^{\lambda_\tau}$  are completely uncertain, since the initial position of each pointer individually is uncertain, however their difference is known, and so  $E_m^{\lambda_\tau} - E_n^{\lambda_0}$  can be obtained. The main shortcoming of this method is that the states  $|q_1 - q_2 = 0, p_1 + p_2 = 0\rangle$  are not in general possible to realise - they are summations over an infinite number of delta functions in the general case of continuous infinite position domain. However some good probe states which are close to this take the form

$$(2.63) \quad |\text{probes}\rangle = \int dq_1 dq_2 e^{\frac{(q_1 - q_2)^2}{4\sigma}} e^{\frac{(q_1 + q_2)^2}{4\Delta}} |q_1\rangle |q_2\rangle$$

up-to normalisation, with  $\Delta \gg \sigma$ . The states in equation (2.63) can be interpreted as source of particles undergoing pair production (and so the momentum ‘conservation’ is given a physical interpretation), where the position of the source has a large uncertainty of  $\Delta$  itself. A realisation of such a measurement is an interesting avenue for further research, perhaps utilising something similar to the GKP states [168].

Bringing the conversation back to the implicit and explicit battery picture, We can describe a quantum measurement in terms of a set of completely positive transformations  $\{C_i\}$  (measurement sub-channels), each associated to a measurement outcome  $i$ . The probability of observing measurement outcome  $i$  when a particular quantum state  $\rho$  is measured as

$$(2.64) \quad p_i = \text{Tr}[C_i[\rho]].$$

The state after the measurement is then

$$(2.65) \quad \rho_i = \frac{C_i[\rho]}{p_i}.$$

We also demand that summation over measurement outcomes must yield a CPTP map, so as to ensure the trace preserving/conservation of probability property;

$$(2.66) \quad C[\rho] = \sum_i \text{Tr}[C_i[\rho]].$$

We will call such a channel the natural channel for a set of measurement sub-channels. As before in the explicit case, we will consider that any energy required for the device to implement the measurement is the same as that which is used to provide the energy for thermodynamic unitaries between the system and bath.

A protocol  $P$  for work extraction using the measurement will be similar to what we have seen in the sections before this, but the resetting operation will in general depend on the measurement result.

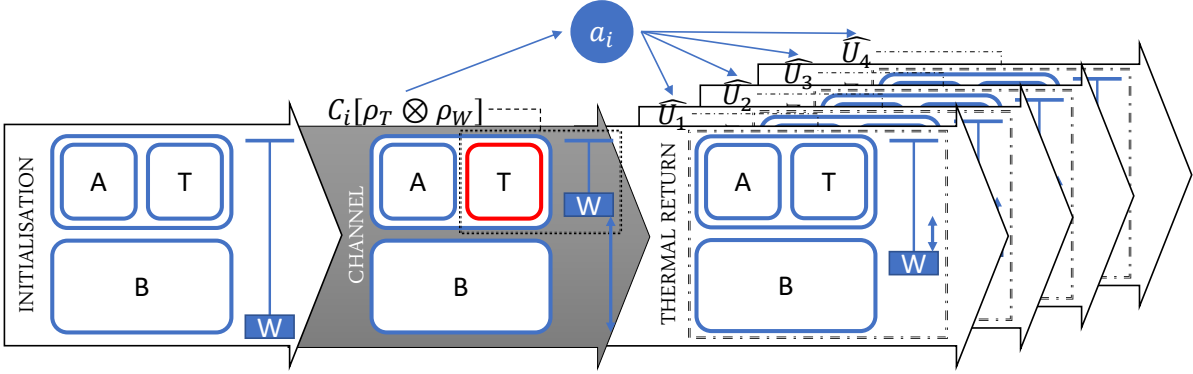


Figure 2.5: Illustration of a single cycle of a thermodynamic protocol  $P$ , which consists of a measurement and a unitary reset. Different unitary resets are applied, dependant upon the measurement outcome obtained in the previous step, to give some feed-forward control over the extractable work.

1. The system and bath are initialised in an uncorrelated state  $\rho = \rho_S \otimes \tau_B$ .
2. The measurement is applied to the target system, giving result  $i$  with probability  $p_i$ , and leaving the system in the state  $\sigma_i = \frac{(C_{s,i}[\rho] \otimes \tau_B)}{p_i}$ , where we define  $C_{s,i}$  as applying  $C_i$  to the target system, and the identity channel to the ancillary system.
3. For each measurement result, a (possibly) different resetting operation is applied that returns the system to its initial state. More specifically we transform  $\sigma_i$  to  $\sigma'_i = \hat{U}_i \sigma_i \hat{U}_i^\dagger$ , where  $\text{Tr}_B[\sigma'_i] = \rho_S$ .

By specifying a protocol  $P$  we are specifying the initial state, the ancillary and bath system, as well as the set of resetting operations  $\{\hat{U}_i\}$ .

The total work extracted by this protocol  $P$  is given by the sum over the work extracted in each step. As before, and in line with the work on metroty, the work gain when applying the measurement must be equal to the energy loss of the system.

**Definition 2.6.** *The work gained on average when making use of measurements  $\{C_i\}$  on the target system  $\rho_T$  (where as before  $\rho_S = \rho_T \otimes \rho_A$ ) is given by*

$$\begin{aligned}
 W_{\{C_i\}}^{\text{apply}} &\equiv - \sum_i p_i \Delta U_{\rho_T - \rho_{T,i}} \\
 &= \sum_i p_i \text{Tr}[H_t \rho_T - H_t \frac{C_i[\rho_T]}{p_i}] \\
 &= \text{Tr}[H_t \rho_T - H_t C(\rho_T)] \\
 (2.67) \qquad &= W_C^{\text{apply}}
 \end{aligned}$$

where  $C[\rho] = \sum_i C_i[\rho]$ .



By making use of definition (2.6), we can calculate the total work benefit provided by a measurement in a thermodynamic cycle as

$$(2.68) \quad W_{\{C_i\},P}^{\text{total}} = W_1 + W_2 + W_3$$

Each step can be either computed or bounded

$$(2.69) \quad \begin{aligned} W_1 &= 0 \\ W_2 &= W_{\{C_i\}}^{\text{apply}} = E(\rho_S) - \sum_i p_i E(\sigma_i) \\ W_3 &\leq \sum_i p_i (F(\sigma_i) - F(\rho_S)). \end{aligned}$$

Hence using the definition of free energy,  $F(\rho) = U(\rho) - TS(\rho)$ , putting everything together leads to,

$$(2.70) \quad \begin{aligned} W_{\{C_i\},P}^{\text{total}} &\leq T \sum_i p_i (S(\rho_S) - S(\sigma_{S,i})) \\ &\leq T \sum_i p_i (S(\rho_T) - S(\sigma_{T,i})) \\ &= T \left( S(\rho_T) - \sum_i p_i S(\sigma_{T,i}) \right) \end{aligned}$$

where  $\sigma_{S,i} = \text{Tr}_B[\sigma_i] = \frac{C_{s,i}[\rho_S]}{p_i}$ , and  $\sigma_{T,i} = \text{Tr}_A[\sigma_{S,i}] = \frac{C_i[\rho_T]}{p_i}$ . As in the case of channels, the last step follows from the monotonicity of the mutual information under the action of local channels.

We note here that for any  $\rho_T$  we can consider protocols in which  $\rho_S = \rho_T$ , meaning that no ancilla is used, and each  $U_i$  is a thermodynamically efficient protocol for transforming  $\sigma_i$  into  $\rho_T$ . Hence, there exist protocols which come as close as required to saturating (2.70). By optimising over protocols to find the work benefit of the measurements, we arrive at the following theorem

**Theorem 2.2.** *The work benefit of the measurement  $\{C_i\}$  is given by*

$$(2.71) \quad W_{\{C_i\}}^{\text{total}} = \sup_P W_{\{C_i\},P}^{\text{total}} = \max_{\rho_T} T(S(\rho_T) - \sum_i p_i S(\sigma_i))$$

As an example, a complete basis measurement [169] has a work benefit of  $T \ln(d_S)$ , taking  $\rho_T$  as the maximally mixed state in the maximisation above, and noting that each post-measurement state is pure and therefore has no entropy. Furthermore, we can now make a robust claim that ‘measurements are of more thermodynamic use than channels’. Of course, in generality the statement is meaningless, and depends on what measurement and what channel, however through the natural channel to measurements, where we recall we associate a channel  $C$  to a collection of measurement sub-channels  $C[\rho] = \sum_i C_i[\rho]$ . By applying the concavity of entropy and theorems (2.2) and (2.1), we find that  $W_C^{\text{total}} \leq W_{\{C_i\}}^{\text{total}}$ , so measurements are capable of providing more work than the corresponding natural channel on average. The reason for this is that measurement and channels, while inducing the same changes in entropy on the state on average,

in the former case we can make use of the measurement information to form an adaptive resetting strategy, which allows us an extra level of control over the thermodynamic cycle.

When considering the work benefit in the context of the second law of thermodynamics, we again find the cost of resetting the measurement device and erasing the information the measurement outcomes carries with itself a thermodynamic cost which outweighs the work benefit. This is of course the solution to the famous paradox of Maxwell's daemon. Let's represent the measurement by a unitary interaction  $\hat{V}$  between the target system and an ancilla stored in the measurement device, denoted by  $Z$ , in initial state  $\rho_Z$ . We have by substitution into the Born rule that for a POVM measurement described by projectors  $\Pi_i$  on the ancilla that

$$(2.72) \quad p_i = \text{Tr}(\hat{V}(\rho_T \otimes \rho_Z)\hat{V}^\dagger(\mathbb{1} \otimes \Pi_i))$$

and

$$(2.73) \quad \sigma_{T,i} = \frac{1}{p_i} \text{Tr}_Z(\hat{V}(\rho_T \otimes \rho_Z)\hat{V}^\dagger(\mathbb{1} \otimes \Pi_i)).$$

The reset unitary on the system, which depends on the measurement result is well described by

$$(2.74) \quad \hat{U} = \sum_i \hat{U}_i \otimes \Pi_i.$$

If the measurement ancilla  $Z$  is reset to its initial state at the end of the protocol, via a resetting interaction with the bath, then we have constructed a fully cyclic protocol on  $T$ ,  $B$  and  $Z$  which returns  $T$  and  $Z$  both to their initial state. We have already shown that such protocols cannot provide a work benefit, so it follows that neither can measurements if the devices that are used for readout must also be reset. For more information on the resource costs of preparing states, and connections with the third law of thermodynamics, we recommend [170]. As a final note, if one includes in the proof an experimenter who reads the measurement device (by interacting with it) then they too must be unitarily reset. Once again, this effect has been investigated in the context of Maxwell daemons [109, 171].

In order to further explore the connections between measurements, channels and explicit batteries, we will consider an interesting scenario, that is deceptive in its simplicity. For a single use of a channel, which transforms  $\rho_T$  to  $\sigma_T$ , we can obtain a work benefit arbitrarily close to  $T(S(\rho_S) - S(\sigma_S))$ . For a measurement the corresponding optimal work benefit is  $\sum_i p_i T(S(\rho_S) - S(\sigma_{S,i}))$ . It is therefore natural to assume that the work benefit upon observing measurement outcome  $i$  is  $T(S(\rho_S) - S(\sigma_{S,i}))$ . However, this intuition is not correct, and the explicit battery case makes it clear why. To show this, let us consider the example of an energy measurement of a qubit with a system Hamiltonian  $\hat{H} = E|1\rangle\langle 1|$ . Assume that the system is initially in a maximally mixed state  $\rho_T = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$ , and that the measurement is characterised by the sub-channels

$$(2.75) \quad \begin{aligned} C_0[\rho] &= |0\rangle\langle 0|\rho|0\rangle\langle 0| \\ C_1[\rho] &= |1\rangle\langle 1|\rho|1\rangle\langle 1|. \end{aligned}$$

We can perform the measurement by preparing a device in the pure 0 state,  $\rho_Z = |0\rangle\langle 0|$  and applying a controlled-NOT operation from the target to the device,

$$(2.76) \quad \hat{V} = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes \hat{X}$$

where  $\hat{X}$  is the Pauli  $x$  gate,  $\hat{X} = |0\rangle\langle 1| + |1\rangle\langle 0|$ . In this case  $[\hat{V}, \hat{H}] = 0$ . Because of this, no interaction with the battery is required to perform the measurement, and hence  $W_{\{C_0, C_1\}}^{\text{apply}} = 0$ . In the case that the measurement result is 0, we know that the state after measurement is  $\sigma_{T,0} = |0\rangle\langle 0|$ , and the optimal work from the reset operation is given by the change in free energy of the reset operation,

$$(2.77) \quad W_{\{C_0, C_1\}0}^{\text{total}} = T \ln 2 - E/2.$$

where we use the notation  $W_{\{C_0, C_1\}0}$  to denote ‘work conditional on result 0 being observed’ in line with the notation used to describe conditional probabilities. If result 1 comes up, then a similar argument gives

$$(2.78) \quad W_{\{C_0, C_1\}1}^{\text{total}} = T \ln 2 + E/2.$$

Each case occurs with probability 1/2, so the overall work benefit is

$$(2.79) \quad W_{\{C_0, C_1\}}^{\text{total}} = T \ln 2.$$

Indeed, this work is equal to  $\sum_i p_i T(S(\rho_S) - S(\sigma_{S,i}))$  as claimed. However, the conditional work benefit in each case (for result  $i$ ) is not equal to  $T(S(\rho_S) - S(\sigma_{S,i})) = T \ln 2$ ! The situation becomes increasingly complicated when the unitaries used to implement the measurement and the Hamiltonian do not commute, and to deal with this case we need to appeal to the explicit battery for guidance.

In order to prove theorem (2.2), we will note that what we must calculate is how much the state of the weight has moved, upon the observation of a measurement outcome. We will refer to this quantity as conditional work benefit. In the explicit battery case, we can show the following theorem.

**Theorem 2.3.** *The conditional work benefit  $W_{\{C_i\}i}^{\text{apply}}$  of applying a measurement  $\{C_i\}$  and obtaining result  $i$ , given that the initial state of the weight is  $|\psi_L\rangle = \frac{1}{\sqrt{2L}} \int_{-L}^L dx |x\rangle$ , and omitting terms of  $O(\frac{1}{L})$  (i.e. taking the large  $L$  limit) is*

$$(2.80) \quad W_{\{C_i\}i}^{\text{apply}} = \frac{1}{p_i} \text{Tr}[C_i \left[ \frac{H_t \rho_T + \rho_T H_t}{2} \right] - H_t C_i[\rho_T]]$$

A proof of this theorem suffices to also prove theorem (2.2) for the case where the state of the weight is  $|\psi_L\rangle = \frac{1}{\sqrt{2L}} \int_{-L}^L dx |x\rangle$ , because by averaging over measurement outcomes

$$(2.81) \quad \begin{aligned} \sum_i p_i W_{\{C_i\}i}^{\text{apply}} &= \text{Tr}[C \left[ \frac{H_t \rho_T + \rho_T H_t}{2} \right] - H_t C[\rho_T]] \\ &= \text{Tr}[H_t \rho_T - H_t C[\rho_T]] \\ &= W_{\{C_i\}}^{\text{apply}}, \end{aligned}$$

we recover the work in the implicit case  $W_{\{C_i\}}^{\text{apply}}$ . Hence, including the work benefit obtained by unitarily resetting the system via thermodynamic operations we find

$$\begin{aligned}
 W_{\{C_i\}i,\rho_S}^{\text{total}} &= W_{\{C_i\}i}^{\text{apply}} + F(\sigma_{S,i}) - F(\rho_S) \\
 &= \frac{1}{p_i} \text{Tr}[C_i \left[ \frac{H_t \rho_T + \rho_T H_t}{2} \right]] - \text{Tr}[H_t \sigma_{T,i}] + F(\sigma_{S,i}) - F(\rho_S) \\
 (2.82) \quad &= \frac{1}{p_i} \text{Tr}[C_i \left[ \frac{H_t \rho_T + \rho_T H_t}{2} \right]] - \text{Tr}[H_t \rho_T] + \text{Tr}[H_A \sigma_{A,i}] - \text{Tr}[H_A \rho_A] + T(S(\rho_S) - S(\sigma_{S,i})).
 \end{aligned}$$

Recall that the system  $S$  is composed of an ancilla  $A$  and the target system  $T$ , there is also an additional ancilla inside the measuring device  $Z$  (which stores the measurement result, but may also have other degrees of freedom), and the ‘weight’  $W$  which stores/provides any work. In the explicit battery formalism, the measurement is carried out by implementing a unitary transformation on  $T$  and  $Z$ , and then performing a projective measurement on  $Z$ .

$$(2.83) \quad C_i[\rho_T] = \text{Tr}_Z[\hat{V} \rho_T \otimes \rho_z \hat{V}^\dagger (I \otimes \Pi_i)],$$

where  $\hat{V}$  is a unitary, and  $\Pi_i$  is a projector onto the outcome space corresponding to result  $i$ . Expanding  $\hat{V}$  in an energy basis as

$$(2.84) \quad \hat{V} = \sum_{abcd} \alpha_{ab}^{cd} |a\rangle \langle b| \otimes |c\rangle \langle d|.$$

We can then extend this to the explicit battery formalism by including a shift operator appropriately;

$$(2.85) \quad \tilde{V} = \sum_{abcd} \alpha_{ab}^{cd} |a\rangle \langle b| \otimes |c\rangle \langle d| \otimes \Gamma_{E_b - E_a}.$$

where  $\Gamma_E$  is a translation operator for the weight. Including the weight gives

$$(2.86) \quad C_i[\rho_T] \simeq \text{Tr}_{ZW}[\tilde{V} \rho_T \otimes \rho_z \otimes \rho_w \tilde{V}^\dagger (I \otimes \Pi_i \otimes I)],$$

where  $\rho_w$  is the initial state of the weight. Considering a ‘top-hat’ wavefunction for the weight,  $\rho_w = |\psi_L\rangle \langle \psi_L|$  where

$$(2.87) \quad |\psi_L\rangle = \frac{1}{\sqrt{2L}} \int_{-L}^L dx |x\rangle,$$

we obtain corrections of  $O(\frac{1}{L})$  which can be made as small as desired by taking  $L$  sufficiently large.

The conditional work benefit of applying the measurement is given by the average of  $\hat{x}$  in the final weight state given that result  $i$  was obtained, minus the initial average of  $\hat{x}$ . As the latter is zero ( $\langle \psi_L | \hat{x} | \psi_L \rangle = 0$ ), we have

$$(2.88) \quad W_{\{C_i\}i}^{\text{apply}} = \frac{1}{p_i} \text{Tr}[(\tilde{V} \rho_T \otimes \rho_z \otimes \tau_w \tilde{V}^\dagger)(\mathbb{1} \otimes \Pi_i \otimes \hat{x})]$$

Substituting in we find that

$$\begin{aligned}
 W_{\{C_i\}i}^{\text{apply}} &= \frac{1}{p_i} \text{Tr} \left( \sum_{abcd} \alpha_{ab}^{cd} |a\rangle \langle b| \otimes |c\rangle \langle d| \otimes \Gamma_{E_b - E_a}(\rho_T \otimes \rho_z \otimes \rho_w) \sum_{efgh} \alpha_{ef}^{gh*} |f\rangle \langle e| \otimes |h\rangle \langle g| \otimes \Gamma_{E_e - E_f}^\dagger(\mathbb{1} \otimes \Pi_i \otimes \hat{x}) \right) \\
 (2.89) \quad &= \frac{1}{p_i} \sum_{abcdefgh} \alpha_{ab}^{cd} \alpha_{ef}^{gh*} \text{Tr} \left( |a\rangle \langle b| \rho_T |f\rangle \langle e| \otimes |c\rangle \langle d| \rho_z |h\rangle \langle g| \Pi_i \otimes \Gamma_{E_b - E_a} \rho_w \Gamma_{E_e - E_f}^\dagger \hat{x} \right).
 \end{aligned}$$

Next, we will demonstrate how to evaluate the trace over the weight system. Consider the general case  $\text{Tr}[\Gamma_A \rho_w \Gamma_B^\dagger \hat{x}]$  where  $A$  and  $B$  are constants. Using the cyclic symmetry of the trace we have

$$\begin{aligned}
 \text{Tr}[\Gamma_A \rho_w \Gamma_B^\dagger \hat{x}] &= \int_{-\infty}^{\infty} dx x \langle x | \Gamma_A \rho_w \Gamma_B^\dagger | x \rangle = \int_{-\infty}^{\infty} dx x \langle x - A | \rho_w | x - B \rangle \\
 &= \int_{-\infty}^{\infty} dx x \psi_L(x - A) \psi_L(x - B) \\
 &= \frac{1}{2L} \int_{\max(A, B) - L}^{\min(A, B) + L} dx x \\
 &= \frac{1}{4L} \left( (\min(A, B) + L)^2 - (\max(A, B) - L)^2 \right) \\
 (2.90) \quad &= \frac{1}{2} (A + B) - \frac{|A^2 - B^2|}{4L}.
 \end{aligned}$$

Considering very large  $L$  for which we can neglect corrections of  $O(\frac{1}{L})$ , we get  $\frac{1}{2}(A + B)$ . Setting  $A = E_f - E_e$  and  $B = E_b - E_a$  and substituting the result into (2.89) we find that  $W_{\{C_i\}i}^{\text{apply}}$  can be given as

$$\begin{aligned}
 W_{\{C_i\}i}^{\text{apply}} &= \frac{1}{p_i} \sum_{abcdefgh} \alpha_{ab}^{cd} \alpha_{ef}^{gh*} \frac{(E_f - E_e) + (E_b - E_a)}{2} \text{Tr} \left( |a\rangle \langle b| \rho_T |f\rangle \langle e| \otimes |c\rangle \langle d| \rho_z |h\rangle \langle g| \right) \\
 &= \frac{1}{p_i} \sum_{abcdefgh} \alpha_{ab}^{cd} \alpha_{ef}^{gh*} \text{Tr} \left( |a\rangle \langle b| \frac{(H_t \rho_T + \rho_T H_t)}{2} |f\rangle \langle e| \otimes |c\rangle \langle d| \rho_z |h\rangle \langle g| \right. \\
 &\quad \left. - H_t |a\rangle \langle b| \rho_T |f\rangle \langle e| \otimes |c\rangle \langle d| \rho_z |h\rangle \langle g| \right) \\
 &= \frac{1}{p_i} \text{Tr} \left( V \frac{(H_t \rho_T + \rho_T H_t)}{2} \otimes \rho_z V^\dagger - H_t V \rho_T \otimes \rho_z V^\dagger \right) \\
 (2.91) \quad &= \frac{1}{p_i} \text{Tr} [C_i \left( \frac{H_t \rho_T + \rho_T H_t}{2} \right) - H_t C_i(\rho_T)]
 \end{aligned}$$

from which we recover (2.3).

There are two important thing to note about the explicit battery formalism for measurements. Firstly, that the size of the ancilla used in the device does not change the results. Secondly, that the state of the weight can and does change. Let us investigate the first issue.

One might be concerned that permitting entanglement between the target measurement system  $T$  and the device ancilla  $A$  would produce work benefits that are dependant on the size of this ancilla. A larger ancilla used to store the measurement result might increase the apparent work benefit of the measurement sub-channel. In fact, it does not matter, and it even does not matter if the state is pure or entangled.

To show this, we note that the system  $S$  is comprised of the target system  $T$  and an ancilla  $A$ . We will show

$$(2.92) \quad S(\rho_S) - \sum_i p_i S(\sigma_{S,i}) \leq S(\rho_T) - \sum_i p_i S(\sigma_{T,i}).$$

Which is to say, a larger system cannot extract more work than the target alone. Here,  $\sigma_{S,i}$  is the post measurement state of the system on obtaining outcome  $i$ , and  $\sigma_{T,i}$  is the post measurement state of the target alone. We can always represent the effect of the measurement by Kraus operators, such that  $\sigma_{S,i} = \frac{1}{p_i} \sum_j (\mathbb{1} \otimes K_{ij}) \rho_S (\mathbb{1} \otimes K_{ij}^\dagger)$  with  $p_i = \sum_j \text{Tr}[K_{ij}^\dagger K_{ij} \rho_T]$  and  $\sigma_{T,i} = \text{Tr}_A[\sigma_{S,i}]$ .

We introduce a third system,  $X$ , which stores measurement results and recalling the non-increasing nature of mutual information under the action of local channels,  $I(A : B) \geq I(A : C[B])$  (which itself implies that  $S(AB) - S(\mathbb{1} \otimes C[AB]) \leq S(B) - S(C[B])$ ). We take  $A$  to be the ancilla, and system  $B$  to be  $\mathcal{H}_T \otimes \mathcal{H}_X$ , the product of the system and storage register, and recalling that our channels are of the form

$$(2.93) \quad C[\rho_{TX}] = \sum_{ij} (K_{ij} \otimes \Gamma_i) \rho_{TX} (K_{ij}^\dagger \otimes \Gamma_i^\dagger),$$

where the  $\Gamma_i$  here are discrete unitary shift operators which act to translate the measurement register by  $i$ ,  $\Gamma_i |0\rangle_X = |i\rangle_X$ . Taking  $\rho_{SX} = \rho_S \otimes |0\rangle\langle 0|$  and substituting all into the monotonicity of mutual information relation for local channels we find that,

$$(2.94) \quad S(\rho_S \otimes |0\rangle\langle 0|) - S\left(\sum_{ij} (\mathbb{1} \otimes K_{ij}) \rho_S (\mathbb{1} \otimes K_{ij}^\dagger) \otimes |i\rangle\langle i|\right) \leq S(\rho_T \otimes |0\rangle\langle 0|) - S\left(\sum_{ij} K_{ij} \rho_T K_{ij}^\dagger \otimes |i\rangle\langle i|\right)$$

and hence

$$(2.95) \quad S(\rho_S) - S\left(\sum_i p_i \sigma_{S,i} \otimes |i\rangle\langle i|\right) \leq S(\rho_T) - S\left(\sum_i p_i \sigma_{T,i} \otimes |i\rangle\langle i|\right).$$

A classically correlated state  $p_i \sigma_i \otimes |i\rangle\langle i|$  has entropy  $S(\sum_i p_i \sigma_i \otimes |i\rangle\langle i|) = \sum_i p_i S(\sigma_i) + H(p_i)$ , where  $H(p_i)$  denotes the Shannon entropy of the distribution specified by  $p_i$ . Thus,

$$(2.96) \quad S(\rho_S) - \sum_i p_i S(\sigma_{S,i}) + H(p_i) \leq S(\rho_T) - \sum_i p_i S(\sigma_{T,i}) + H(p_i)$$

which completes the proof that having access to a larger system size can not effect the work associated to the measurement channel.

The second issue is that of weight state dependence. In order that thermodynamic operations, a channel, or measurement implements correctly it suffices that the state of the weight is sufficiently peaked about the zero of momentum. However, the conditional work extraction is different in this regard. In this section, we will show that the state of the weight, can alter the value of the conditional work extraction. In particular, there is no longer universal bound in terms of  $\epsilon$ 's and  $\delta$ 's that ensures we implement the desired transformation from the implicit battery

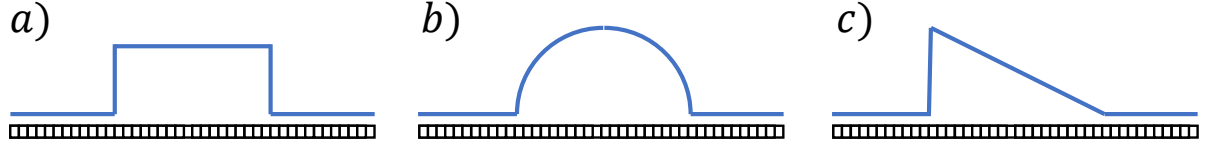


Figure 2.6: Three possible weight states. In a) we depict the top hat weight state, in b) we depict a real symmetric weight state. In c) we depict a weight state that is real, but not symmetric.

case in the explicit case. This is somewhat expected, it is difficult to even define a conditional work extraction for the implicit case, however it is also not without interest that thermodynamic quantities depend upon the features of the machines used to implement them. Up until now, we have used a broad ‘top-hat’ function  $|\psi_L\rangle = \frac{1}{\sqrt{2L}} \int_{-L}^L |x\rangle dx$  for the initial state of the weight. However, one can easily come up with alternative, and other reasonable initial states. In fact, a broad class of reasonable weight states do leave the conditional work extraction as defined in theorem (2.3) invariant. This class is the real symmetric weight states, which satisfy

$$(2.97) \quad \int_{-\infty}^{\infty} dx \phi_L(x-C)\phi_L(x+C) = 1 + O\left(\frac{1}{L}\right).$$

One can see this because from the large  $L$  limit which we employ,

$$(2.98) \quad \begin{aligned} \text{Tr} \Gamma_A \rho_w \Gamma_B^\dagger \hat{x} &= \int_{-\infty}^{\infty} dx x \phi_L(x-A)\phi_L(x-B) = \int_{-\infty}^{\infty} dx \left(x + \frac{A+B}{2}\right) \phi_L\left(x - \frac{A-B}{2}\right) \phi_L\left(x + \frac{A-B}{2}\right) \\ &= \frac{A+B}{2} \int_{-\infty}^{\infty} dx \phi_L\left(x - \frac{A-B}{2}\right) \phi_L\left(x + \frac{A-B}{2}\right) \\ &= \frac{A+B}{2} + O\left(\frac{1}{L}\right), \end{aligned}$$

where in the third line we have used the symmetry of  $\phi_L(x)$ . This means that a physically reasonable class of weight states such as the Gaussian will also leave the conditional work extraction invariant. Let us consider what happens when outside of this class. We will consider an anti-symmetric triangular weight state:

$$(2.99) \quad |\tau_w\rangle = \sqrt{\frac{3}{L^3}} \int_{-\frac{3L}{4}}^{\frac{L}{4}} \left(x + \frac{3L}{4}\right) |x\rangle dx,$$

where the constants are chosen so that the state is normalised, and  $\langle \tau_L | \hat{x} | \tau_L \rangle = 0$ . Note that this state still has a narrow momentum distribution about  $p = 0$  in the limit of large  $L$ , in keeping with our other results. Taking  $\rho_w = |\tau_L\rangle \langle \tau_L|$  we have

$$(2.100) \quad \begin{aligned} \text{Tr}[\Gamma_A \rho_w \Gamma_B^\dagger \hat{x}] &= \int_{-\infty}^{\infty} dx x \tau_L(x-A)\tau_L(x-B) = \frac{3}{L^3} \int_{\max(A,B)-\frac{3L}{4}}^{\min(A,B)+\frac{L}{4}} dx x \left(x + \frac{3L}{4} - A\right) \left(x + \frac{3L}{4} - B\right) \\ &= \frac{A+B}{8} + \frac{3 \min(A,B)}{4} + O\left(\frac{1}{L}\right) \\ &= \frac{A+B}{2} - \frac{3|A-B|}{8} + O\left(\frac{1}{L}\right) \end{aligned}$$

Note that the first term is the same as we got for the ‘top-hat’ weight state. To calculate the effect of the second term, it is helpful to express the measurement using a Kraus decomposition.

$$(2.101) \quad C_i[\rho_T] = \sum_j K_{ij} \rho_T K_{ij}^\dagger.$$

Extending this to incorporate an explicit weight, we define

$$(2.102) \quad \tilde{K}_{ij} = \sum_{a,b} (K_{ij})_{ab} |a\rangle \langle b| \otimes \Gamma_{E_b - E_a}.$$

Then

$$\begin{aligned} W_{\{C_i\}i}^{\text{apply}} &= \sum_j \frac{1}{p_i} \text{Tr}[(\tilde{K}_{ij} \rho_T \otimes \tau_w \tilde{K}_{ij}^\dagger)(\mathbb{1} \otimes \hat{x})] \\ &= \frac{1}{p_i} \sum_{jabcd} (K_{ij})_{ab} (K_{ij}^\dagger)_{cd} \text{Tr}[|a\rangle \langle b| \rho_T |c\rangle \langle d| \otimes \Gamma_{E_b - E_a} \rho_w \Gamma_{E_c - E_d}^\dagger \hat{x}] \\ &= \frac{1}{p_i} \sum_{jabcd} (K_{ij})_{ab} (K_{ij}^\dagger)_{cd} \text{Tr}\left[|a\rangle \langle b| \frac{H_t \rho_T + \rho_T H_t}{2} |c\rangle \langle d| - H_t |a\rangle \langle b| \rho_T |c\rangle \langle d| \right. \\ &\quad \left. - \frac{3}{8} |E_b - E_a - E_c + E_d| |a\rangle \langle b| \rho_T |c\rangle \langle d| \right] \\ &= \text{Tr}\left[C_i \left[ \frac{H_t \rho_T + \rho_T H_t}{2} \right]\right] - \text{Tr}[H_t C_i[\rho_T]] - \frac{3}{8 p_i} \sum_{jabc} (K_{ij}^\dagger)_{ca} (K_{ij})_{ab} |E_b - E_c| \langle b| \rho_T |c\rangle \\ &= \text{Tr}\left[C_i \left[ \frac{H_t \rho_T + \rho_T H_t}{2} \right]\right] - \text{Tr}[H_t C_i[\rho_T]] - \frac{3}{8 p_i} \sum_{bc} (M_i)_{cb} |E_b - E_c| \langle b| \rho_T |c\rangle \end{aligned}$$

where  $M_i = \sum_j K_{ij}^\dagger K_{ij}$  is the POVM element corresponding to result  $i$  (which satisfies  $p_i = \text{Tr}[M_i \rho_T]$ ), a quantity which is independent of the choice of Kraus decomposition.

Note that the extra term above compared to our previous result is dependant on the off-diagonal elements of  $\rho$  and the measurement operator  $M$  in its energy eigenbasis. This state of the weight produces a different energy shift to the top-hat state when the state of the system contain energy coherence’s, even when  $L \rightarrow \infty$ . Simply put - that the conditional work extraction when applying a measurement is not operationally well defined without specifying the weight state in detail. Furthermore, effects of this type are of a genuinely quantum nature and appear into our description of thermodynamics.

It is not clear how large the set of weight states is, which lead to an additional work benefit term compared to the top-hat weight state. Perhaps there are only a few ‘bad’ weight states which lead to this (e.g. those containing both spatial non-continuity and asymmetry), in which case the result of theorem (2.3) could reasonably be taken as standard. This is an interesting avenue for further investigation.

This completes our discussion of theorem 2.2 and theorem 2.3. Note that the explicit battery case has allowed us to identify both

1. A characterisation of the work like energy which occurs for the observation of each measurement outcome observed when making a POVM.



2. Compatibility with the second law and quantum measurement.

At the cost of,

1. Dependence on the mechanism which is used to make the measurement, i.e. a weight state dependence. By making use of the real symmetric weight states, we can formulate a consistent picture.
2. The second law as a statistical phenomenon, not an absolute physical law. In other words, the second law holds on the average <sup>1</sup>.

## 2.5 Post-Selection

Given our characterisation of the conditional work benefit in theorem (2.3), we find we can naturally extend this to the case of post-selected measurements.

As in the introduction, post-selection is the action of discarding statistics from a data set. In the context of quantum theory, post-selection eludes to many foundational questions for physics [1], and also has practical importance in the context of taking weak measurements [173], and quantum algorithms [174, 175].

A measurement is well described by a set of POVM element's, (or sub-channels  $\{C_i\}$ ), which sum to the identity (completely positive maps that sum to preserve the trace). If we denote some arbitrary subset of measurement outcomes as successes  $i \in succ$ , and the rest as fails,  $i \in fail$ , then post-selected measurements are collections of measurement sub-channels where the outcomes lie only within the *succ* subset. To achieve this, the simplest way is to make a regular quantum measurement and then discard the statistics corresponding to outcomes in *fail*. It is convenient to define the probability of post-selection success,  $p_{succ} = \sum_{i \in succ} p_i$ , where  $p_i = \text{Tr}[C_i[\rho]]$ . By making use of the work benefit in the previous section we can calculate the conditional work benefit of a post-selected measurement by,

$$(2.103) \quad W_{\{C_i\}|succ,\rho_S}^{\text{total}} = \frac{1}{p_{succ}} \sum_{i \in succ} p_i W_{\{C_i\}|i,\rho_S}^{\text{total}}.$$

Placing ourselves within the class of real symmetric weight states, we can calculate an alternative formula for  $W_{\{C_i\}|succ,\rho_S}^{\text{total}}$ . Making use of efficient, adaptive thermodynamic transformations for the reset we obtain

$$(2.104) \quad W_{\{C_i\}|succ,\rho_S}^{\text{total}} = \sum_{i \in succ} \text{Tr}[C_i \left[ \frac{H_T \rho_T + \rho_T H_T}{2 p_{succ}} \right]] - \text{Tr}[H_T \rho_T] \\ + \sum_{i \in succ} \frac{p_i}{p_{succ}} \text{Tr}[H_A \sigma_{A,i}] - \text{Tr}[H_A \rho_A] + T \left( S(\rho_S) - \sum_{i \in succ} \frac{p_i}{p_{succ}} S(\sigma_{S,i}) \right).$$

---

<sup>1</sup>Actually, there are rather good arguments that another classical phenomena, conservation laws, are also statistical in quantum theory [172]

Considering the binary case, where the measurement has only two outcomes and so is described by the set of two completely positive operators  $\{C_{\text{succ}}, C_{\text{fail}}\}$  we find

$$(2.105) \quad W_{\{C_{\text{succ}}, C_{\text{fail}}\}|\text{succ}, \rho_S}^{\text{total}} = \text{Tr}[C_{\text{succ}} \left[ \frac{H_T \rho_T + \rho_T H_T}{2 p_{\text{succ}}} \right]] - \text{Tr}[H_T \rho_T] + \text{Tr}[H_A \sigma_{A, \text{succ}}] - \text{Tr}[H_A \rho_A] + T(S(\rho_S) - S(\sigma_{S, \text{succ}})).$$

Post-selection can be thought of as a binary filtering of measurement outcomes into keep and discard. Moreover, any post-selected measurement can be coarse grained into the success and fail categories, we do not define any notion of partial post-selection here for simplicity. In this case, one can see by comparing (2.104) and (2.105) and using the concavity of the entropy that  $W_{\{C_i\}|\text{succ}, \rho_S}^{\text{total}} \geq W_{\{C_{\text{succ}}, C_{\text{fail}}\}|\text{succ}, \rho_S}^{\text{total}}$ , so in some sense the binary post-selection is the minimal example that captures the power of post-selection for thermodynamics while omitting considerations surrounding the resolution of the post-selected measurements.

The post-selected formalism is surprising in many ways, and perhaps the most is that any non-trivial post-selected channel can lead to an unbounded work benefit. Moreover, this is not a property of many uses of the post-selected channel, a single use is sufficient to allow this. This rests on the (now broken) independence of the measurement upon the system size.

**Theorem 2.4.** *Any non-trivial post-selected measurement (i.e. where the success probability has some dependence on the state) can be used to obtain an unbounded work benefit.*

$$(2.106) \quad W_{\{C_i\}|\text{succ}}^{\text{total}} = \sup_{\rho_S} W_{\{C_i\}|\text{succ}, \rho_S}^{\text{total}} = \infty.$$

To illustrate this, we will begin with a simple example. Consider a qubit in state  $\rho_T$ , and a post-selected measurement where

$$(2.107) \quad \begin{aligned} C_{\text{succ}}[\rho_T] &= |0\rangle \langle 0| \rho_T |0\rangle \langle 0| \\ C_{\text{fail}}[\rho_T] &= |1\rangle \langle 1| \rho_T |1\rangle \langle 1|. \end{aligned}$$

and let the system have a trivial Hamiltonian,  $\hat{H}_S = 0$ . Taking the initial state of the system (in  $\mathcal{H}_T \otimes \mathcal{H}_A$ ) to be

$$(2.108) \quad \rho_S = \frac{1}{2} |0\rangle \langle 0| \otimes |\phi\rangle \langle \phi| + \frac{1}{2} |1\rangle \langle 1| \otimes \frac{1}{d_A} \mathbb{1},$$

where  $|\phi\rangle$  is any pure state of the ancilla, and  $d_A$  is the ancilla dimension. Upon a successful post-selection, the state collapses to  $\sigma_{S, \text{succ}} = |0\rangle \langle 0| \otimes |\phi\rangle \langle \phi|$ . This state is pure, and has zero entropy. By using efficient thermal operations on the system and bath to take  $\sigma_{S, \text{succ}}$  back to  $\rho_S$ , we can extract work  $W = F(\sigma_{S, \text{succ}}) - F(\rho_S) = T(S(\rho_S) - S(\sigma_{S, \text{succ}})) = T(\ln 2 + \frac{1}{2} \ln d_A)$ . This work grows with the dimension of the ancilla  $d_A$  as  $\ln(d_A)$ , which is unbounded.

We can prove this statement in generality as follows. Consider a post-selected measurement  $\{C_{\text{succ}}, C_{\text{fail}}\}$ . Given any completely positive map  $C_{\text{succ}}$ , there exists a positive operator  $M_{\text{succ}}$

such that  $p_{\text{succ}} = \text{Tr}[C_{\text{succ}}[\rho_T]] = \text{Tr}[M_{\text{succ}}\rho_T]^2$ . Let  $|u\rangle$  and  $|v\rangle$  be eigenvectors of  $M_{\text{succ}}$  with the maximal and minimal eigenvalue respectively,

$$(2.109) \quad M_{\text{succ}}|u\rangle = \lambda_{\text{max}}|u\rangle,$$

$$(2.110) \quad M_{\text{succ}}|v\rangle = \lambda_{\text{min}}|v\rangle.$$

The restriction that the post-selection is not trivial means that  $\lambda_{\text{max}} > \lambda_{\text{min}}$ . For any such trivial measurement,  $C_{\text{succ}} = \alpha D$  where  $\alpha$  is a constant and  $D$  is a trace-preserving channel. This corresponds to performing  $D$  and then failing independently at random with probability  $(1 - \alpha)$ , which for all practical purposes is the same as just performing  $D$  fewer times. There will be more on this concept in chapter 4. Proceeding with our proof, application of  $C$  to the state

$$(2.111) \quad \rho_S = \frac{1}{2}|u\rangle\langle u|_T \otimes |\phi\rangle\langle\phi|_A + \frac{1}{2}|v\rangle\langle v|_T \otimes \frac{1}{d_A}\mathbb{1}_A,$$

where the Hamiltonian for the ancilla is trivial leads to a work benefit, by substitution into equation (2.105),

$$(2.112) \quad W_{\{C_{\text{succ}}, C_{\text{fail}}\}|\text{succ}, \rho_S}^{\text{total}} = \text{Tr}\left[L_{\text{succ}}\left(\frac{H_t\rho_T + \rho_T H_t}{2p_{\text{succ}}}\right)\right] - \text{Tr}[H_t\rho_T] + \text{Tr}[H_A\sigma_{A,\text{succ}}] - \text{Tr}[H_A\rho_A] + T(S(\rho_S) - S(\sigma_{S,\text{succ}}))$$

By using the cyclic symmetry of the trace, the definitions of  $|u\rangle$  and  $|v\rangle$ , and the fact that  $p_{\text{succ}} = (\lambda_{\text{max}} + \lambda_{\text{min}})/2$  to simplify the first term on the right-hand side. Also using  $H_A = 0$  and  $S(\rho_S) = \ln 2 + \frac{1}{2}\ln d_A$  gives

$$(2.113) \quad W_{\{C_{\text{succ}}, C_{\text{fail}}\}|\text{succ}, \rho_S}^{\text{total}} = \frac{\lambda_{\text{max}}\langle u|H_t|u\rangle + \lambda_{\text{min}}\langle v|H_t|v\rangle}{\lambda_{\text{max}} + \lambda_{\text{min}}} - \frac{\langle u|H_t|u\rangle + \langle v|H_t|v\rangle}{2} + T\left(\ln 2 + \frac{1}{2}\ln d_A - S(\sigma_{S,\text{succ}})\right)$$

Note that  $\frac{\lambda_{\text{max}}\langle u|H_t|u\rangle + \lambda_{\text{min}}\langle v|H_t|v\rangle}{\lambda_{\text{max}} + \lambda_{\text{min}}}$  and  $\frac{\langle u|H_t|u\rangle + \langle v|H_t|v\rangle}{2}$  both represent convex mixtures of energies. this means that they lie within the largest and smallest eigenvalues of  $H_T$ , which we denote respectively by  $E_{\text{max}}$  and  $E_{\text{min}}$ . Hence

$$(2.114) \quad W_{\{C_{\text{succ}}, C_{\text{fail}}\}|\text{succ}, \rho_S}^{\text{total}} \geq E_{\text{min}} - E_{\text{max}} + T\left(\ln 2 + \frac{1}{2}\ln d_A - S(\sigma_{S,\text{succ}})\right)$$

Furthermore, from the subadditivity of the entropy we have

$$(2.115) \quad S(\sigma_{S,\text{succ}}) \leq S(\sigma_{T,\text{succ}}) + S(\sigma_{A,\text{succ}}) \leq \ln(d_t) + S(\sigma_{A,\text{succ}}).$$

Defining

$$(2.116) \quad q = \frac{\lambda_{\text{min}}}{\lambda_{\text{max}} + \lambda_{\text{min}}},$$

<sup>2</sup>Writing  $C_{\text{succ}}$  in Kraus decomposition as  $C_{\text{succ}}[\rho] = \sum_j K_j \rho K_j^\dagger$ , we have  $M_{\text{succ}} = \sum_j K_j^\dagger K_j$

note that

$$(2.117) \quad \sigma_{A,\text{succ}} = (1-q)|\phi\rangle\langle\phi| + q\left(\frac{1}{d_A}\mathbb{1}_A\right)$$

Using the fact that  $S(\sum_i p_i \rho_i) \leq H(p_i) + \sum_i p_i S(\rho_i)$ , where  $H(p_i)$  is the Shannon entropy, we have

$$(2.118) \quad S(\sigma_{A,\text{succ}}) \leq H(q) + q \ln d_A \leq \ln 2 + q \ln d_A$$

Putting everything together we then have

$$(2.119) \quad W_{\{C_{\text{succ}}, C_{\text{fail}}\}|\text{succ}, \rho_S}^{\text{total}} \geq E_{\text{min}} - E_{\text{max}} + T \left( \left( \frac{1}{2} - q \right) \ln d_A - \ln d_t \right).$$

As  $\lambda_{\text{max}} > \lambda_{\text{min}}$ , we have  $q < \frac{1}{2}$ . Hence the work benefit grows with  $\ln d_A$ .

This result was derived for a real symmetric state of the weight, namely the top-hat weight state and does not include  $O(1/L)$  corrections due to the finite width of that state, and  $O(\epsilon)$  corrections due to inefficiencies in the thermodynamic protocol for resetting the system state. Both of these can be made as small as desired, and so do not effect the logarithmic divergence of the work with the dimension of the ancilla system. Using other weight states that are sharply peaked around momentum zero would also not change the scaling with  $d_A$ .

## 2.6 Discussions and Conclusions

We have laid out a framework for thermodynamics in quantum theory. In it, we have shown how unitary operations provide thermodynamic efficient transformations of a state, able to extract work at the optimum amount as dictated by the second law - i.e. extract work equal to the free energy change in the state that is induced by the transformation. We have shown that adding an explicit model of a quantum battery does not interfere with this upper bound, at least in the ideal case. We have given and justified a quantifier for the thermodynamic utility of a quantum channel, and subsequently shown some interesting relations between a variety of types of quantum channels. We have given a thermodynamic picture of measurement, which is novel as many resource theoretic approaches to thermodynamics omit the measurement problem totally. After this, we gave a formula to calculate the conditional work extraction - which represents how much we expect the state of a weight system to have moved conditional on observing certain measurement outcomes. Finally, we investigated the thermodynamics of post-selected systems, and shown how a post-selected channel can lead to an unbounded amount of work - for any amount of post-selection probability.

The physical principle in this chapter is energy conservation. Without some form of energy conservation, all of our work doesn't matter in the slightest.



## CAUSALITY IS CLASSICAL IN QUANTUM THEORY

This chapter is based off of the paper ‘Quantum theory cannot violate a causal inequality’ [82] which has been published in PRL. This work was completed at the University of Bristol in conjunction with my supervisor, Dr. Anthony J Short. In it, we present a novel characterisation of quantum theory with coherent control over laboratory orderings in a causal scenarios, and show that it cannot be used to violate causal inequalities.

A very similar result was obtained by Wechs *et al.* [176], which was posted on the ArXiv on the very same day as our paper. This result however was obtained in a very different manner, and appeals extensively to the process matrix formalism.

### 3.1 Introduction

Quantum theory permits superpositions. In the famous two slit experiment [177], where particles are scattered through a grating with two (or more) gaps, one finds that the particles behave like waves. More specifically, when detected at a screen some distance from the grating the particle diffraction pattern resembles a wave-like interference. Importantly, this interference pattern has a characteristic sinusoidal quality, where there are some locations that are more likely for the particles to be found than others - even when the particles are sent through the grating one at a time. One must come to the conclusion that the particle has behaved like a wave, and so the answer to the question “which slit did the particle pass through?”, is “both!”. Stranger still, attempt to perform a measurement upon the which slit information degrades the interference pattern, and destroys it totally when one is sure through which slit the particle has travelled. We would say that the which path information can be placed into superposition.

Quantum theory also permits entanglement. Entanglement is a special type of correlation

between systems, where a bipartite system cannot be decomposed into a mixture of states that are pure for either party. It is, in some sense meaningless or incomplete to consider the state of a single system, they are a single object.

The question remains then, if some physical properties can be placed into superposition or entangled, which physical properties can be certified to be in a superposition or an entangled state?

In the example of entanglement, it is possible to perform a Bell test to certify the quantum properties of the correlations between systems. The violation of a Bell inequality is a hallmark of a genuinely quantum type of correlation between the systems. Properties such as spin of a fermion or polarisation of a photon can be entangled.

In this chapter, we will investigate certifiable violations of a different type of inequality - a causal inequality - and the property wherein the causal order of events is placed into a superposition.

When describing a physical process, a common underlying assumption across most of the field is that the causal order of events can be ascertained. Earlier events can effect later ones, and later events cannot effect earlier ones. With the exception of closed-time-like-curves in general relativity, descriptions of events that contradict this assumption are left to the big screens of Hollywood in movies such as 'Back To The Future' or in television such as 'Dr Who'. Recently, this hard boundary has been challenged, applying concepts such as superposition or the uncertainty principle to the causal order of events leads one to 'causal indefiniteness' [178, 179, 180, 181, 182]. This phenomenon has been tested experimentally [183, 184, 185], can be exploited to give surprising advantages in certain tasks [178, 179, 186, 187]. Causal indefiniteness can help distinguishing between unitary pairs that commute or anti-commute, or to combine channels which individually have a zero capacity to an aggregate object with a non-zero capacity. An interesting question is whether quantum theory can generate results which cannot be simulated by any model with a classically causal structure. Such results would violate a causal inequality [188, 189, 62, 190].

In this chapter, we focus on the relationship between the type of causal indefiniteness present in quantum theory and the power needed to violate causal inequalities. We will show that, despite certainly allowing causally indefinite processes, quantum theory cannot violate causal inequalities. Hence, quantum theory cannot be used to gain advantage over classically causal process for tasks defined in a theory independent manner (such as the 'guess your neighbours input tasks' in the introduction chapter). Previous works have shown that the 'quantum switch' scenarios cannot be used to violate causal inequalities [191], and that causal order cannot be placed into a pure superposition [192]. It has also been shown the causal inequality violations are possible when we condition on measurement outcomes of one party [193]. This notion of post-selection has been shown to implement a 'superposition of causal orders' [53, 194, 195, 196]. However, our results in this chapter show that such violations are not possible without

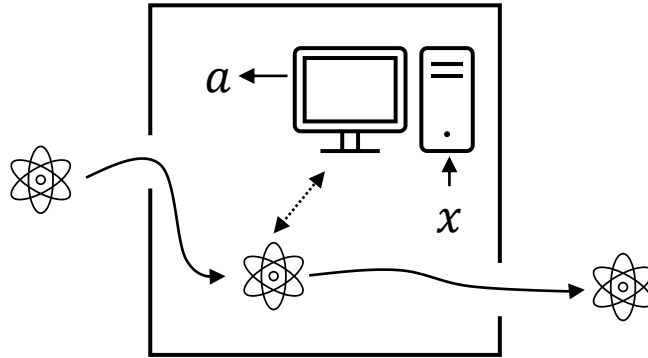


Figure 3.1: Simple representation of Alice's local laboratory, during an experiment. At some point, a system enters her lab. A measurement is made of that system which returns outcome  $a$ , conditional on some variable  $x$ . The system is then permitted to exit her lab.

conditioning.

### 3.2 Causal Process

Let us first refresh our discourse in the introduction surrounding causality in the light of quantum theory. We will first set up a paradigm in which it is possible to prove statements about causality in quantum theory. We will consider a set of parties, who act within the paradigm of local laboratories to conduct experiments.

It is easiest to consider a simple case first. Let us consider that there are only two parties in the entire experiment, Alice and Bob. In the spirit of general probabilistic theories [197] Alice will be provided measurement settings  $x$  and produce outcomes  $a$ . Bob will be provided measurement settings  $y$  and provide outcomes  $b$ . During the experiment, a system will enter their laboratory once and only once (and they will be aware of this fact!). They will perform measurements upon this system and record the outcomes of those measurements in a result register. They will also perform transformations upon the system - that is, change the state of the system. Note that measurements in quantum mechanics also change the state of the system. Apart from through the system, the parties do not communicate with one another.

It is simple to imagine, as an illustrative aside, a classical example. In this instance, the system could be a coin, with states  $H$  (standing for heads) and  $T$  (standing for tails). The measurement procedure could be to record the state of the coin as it comes into the lab. The transformation procedure could be to perform a coin flip, that is, randomise the state of the coin. An alternative transformation could be to always set the coin to  $H$  if the input is 0, and set to  $T$  if the input is 1. An alternative still could be just to always set the coin to  $H$  when you pass it outside the laboratory, ignoring the input totally. Even the process of measurement can display tricksiness in this simple example - Alice may be following a procedure where if the coin is given to her in the state  $T$ , she will record  $H$ .



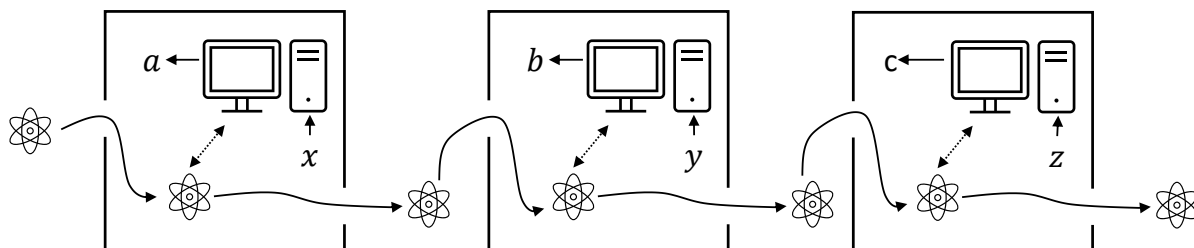


Figure 3.2: A simple causal process, where Alice is first, Bob is second and Charlie is third in the causal order. This process is deterministic, in the sense that the causal order is not associated to some probability which is  $< 1$ . What goes on inside the labs however, might well be probabilistic.

What ever the procedure, the results of all such experiments will be able to be well described by the conditional probability  $p(a, b|x, y)$ . However, not all conditional probabilities  $p(a, b|x, y)$  can be achieved through means for a causal process. Examples of distributions which cannot be achieved by this causal paradigm we have seen in equation (1.61), and equation (1.62).

In this example, the most general causal process would be to choose randomly who will receive the system first (with probability  $p(\text{Alice first})$ ,  $p(\text{Bob first})$ ). If Alice is first, her measurement result could depend upon her measurement setting, she has access to that information, but it cannot depend on the measurement setting or outcome of Bob - these have not yet occurred. In this case then, her marginal probability is well described by  $p(a|x)$ . As each party must see the system exactly once, in the case where Alice has acted first, then we have a completeness relation  $p(\text{Bob second}|\text{Alice First}) = 1$ . Thus, the system enters Bob's laboratory, with his marginal now potentially dependant on everything that has happened thus far,  $p(b|y, a, x)$ . Considering the other causal order means for this bipartite example, and mixing of the causal order based on classical randomness we obtain

$$(3.1) \quad p^{\text{causal}}(a, b|x, y) = p(\text{Alice first})p(a|x)p(b|a, x, y) + p(\text{Bob first})p(b|y)p(a|b, x, y).$$

If we use appropriate notation, the case of multiple parties is a straightforward generalisation<sup>1</sup>. Observe the above causal probability contains two types of terms. The first, such as  $p(\text{Alice first})$ ,  $p(\text{Bob first})$  and  $p(\text{Bob second}|\text{Alice first})$  determine the causal order in which the parties act. and the second, such as  $p(a|x)$ ,  $p(b|y, a, x)$ ,  $p(b|y)$  and  $p(a|b, x, y)$  determine the outcome probabilities of measurements conditioned on the specific measurement and transformation procedure used. We will extend this to multiple parties by considering the 'who is next?' and 'what did they see?' probabilities independently. We use  $l_k$  to denote the  $k$ th party that receives the system, and denote the probability for this to occur by  $p_k(l_k|H_{k-1})$ . The conditional on  $H_{k-1}$  represents the history (including all previous parties that have measured, and their inputs and outputs) for it should be permitted for parties in the causal past of  $l_k$  to affect who is the next party to act. Note that since this is the bipartite case, we did not need any conditional on the first

<sup>1</sup>That is not to say that there is anything trivial about moving between bipartite and tripartite experiments in terms of proving anything, or the inherent power in each scenario. For example, the results in [198] and [64].

parties output and input as we have a completeness relation, i.e.  $p(\text{Bob second}|\text{Alice First}, a, x) = p(\text{Bob second}|\text{Alice First}) = 1$ .

Consider now a tripartite experiment, with Alice, Bob and Charlie as the participants. In the case that Charlie comes first, the system could be passed to Alice or Bob next, and moreover may be passed based on the outcome of his measurement. Here,  $p_2(\text{Alice next}|\text{Charlie got outcome} = 1)$  may not be equal to  $p_2(\text{Alice next}|\text{Charlie got outcome} = 0)$ . The probability  $p_k(l_k|H_{k-1})$  accounts for these scenarios [190, 198].

The probability for  $l_k$  to obtain given results can depend on this history, not on the future, and of course on the measurement setting, denoted  $x_{l_k}$ . We write this probability as  $p_k(a_{l_k}|H_{k-1}, x_{l_k})$ . A causal model is then nothing more than the summation over all available parties at all stages of the measurement procedure.

**Definition 3.1.** *A causal probabilistic model can be written as*

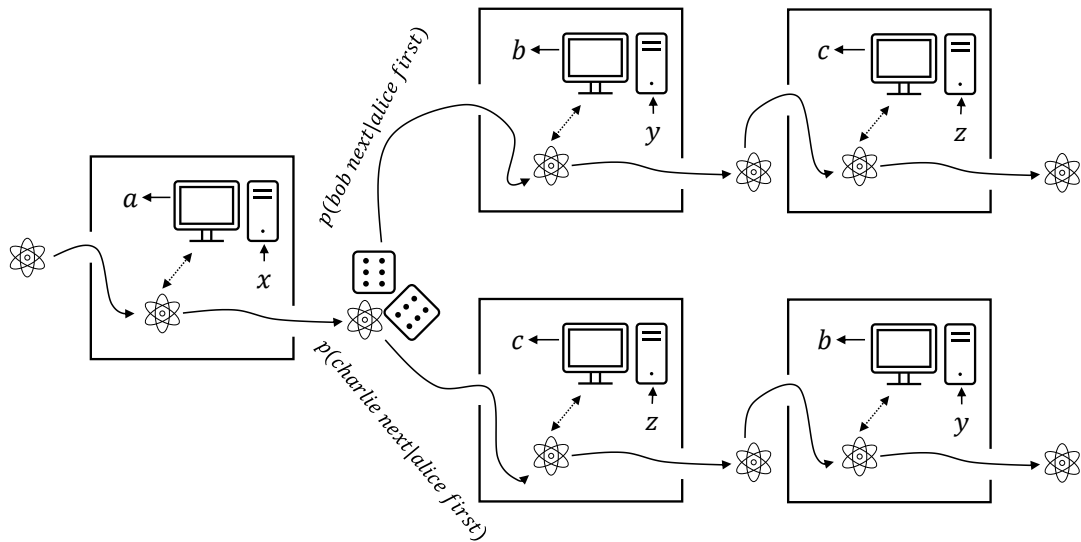
$$(3.2) \quad p^{\text{causal}}(\vec{a}|\vec{x}) = \sum_{l_1 \notin \mathcal{L}_0} \dots \sum_{l_N \notin \mathcal{L}_{N-1}} p_1(l_1|H_0) p_1(a_{l_1}|H_0, x_{l_1}) p_N(l_N|H_{N-1}) p_N(a_{l_N}|H_{N-1}, x_{l_N})$$

where the  $p_k(l_k|H_{k-1})$  terms represent probabilities for party  $l_k$  to act at stage  $k$  of the causal order, and  $p_k(a_{l_k}|H_{k-1}, x_{l_k})$  terms represent probabilities for party  $l_k$ , who has acted at stage  $k$  of the causal order to obtain measurement result  $a_{l_k}$ . Both of the above probabilities are conditional on a history,  $H_{k-1}$ , which contains all of the information about previous inputs, outputs and party order. In particular, the history  $H_k = (h_1, \dots, h_k)$  is the ordered list of triples  $h_i = (l_i, a_{l_i}, x_{l_i})$ . The summations are performed over all possible next parties, excluding parties who have already acted, which are stored in the unordered sets  $\mathcal{L}_k = \{l_1, \dots, l_k\}$ . To emphasise the symmetry between the terms we include  $H_0$  and  $\mathcal{L}_0$ , which are defined as empty sets, as no parties have acted at that point.

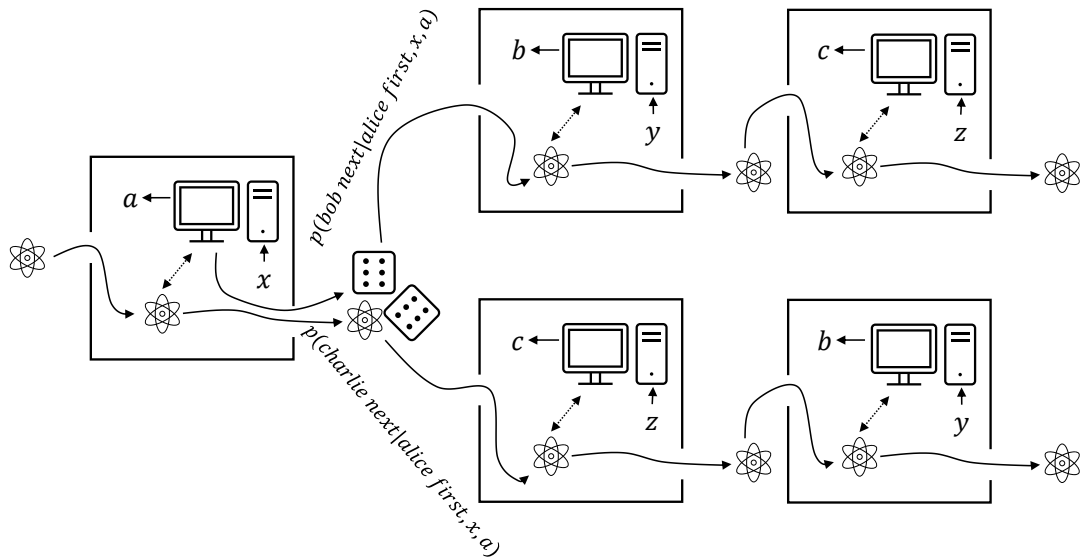
Note that this definition (3.1) defines a convex set of probabilities, since the convex combination of two such causal probabilities gives a causal probability;

$$(3.3) \quad \alpha p_1^{\text{causal}}(\vec{a}|\vec{x}) + (1 - \alpha) p_2^{\text{causal}}(\vec{a}|\vec{x}) = p^{\text{causal}}(\vec{a}|\vec{x}).$$

This is not a trivial statement however a proof can be found in [199]. Linear constraints on this convex set which are satisfied by every  $p^{\text{causal}}(\vec{a}|\vec{x})$ , but can be violated by a general  $p(\vec{a}|\vec{x})$  are known as ‘causal inequalities’, a temporal analogue of Bell inequalities [29], which have been widely studied in the literature in the context of non-locality. Thus, by definition any  $p^{\text{causal}}(\vec{a}|\vec{x})$  cannot violate a causal inequality. A violation of a causal inequality in nature, by either observation in experiment or calculation in some physical theory, would necessarily have radical consequences for our interpretations of physics. Their arrival would signify that there are some phenomena that do not have a causal grounding, in a similar way to which Einstein considered events to lack a grounding in a single global time variable.



(a) This causal process is probabilistic determined. After Alice has acted, some dice are rolled and the other two options for the causal order (Bob, then Charlie / Charlie then Bob) are decided between.



(b) As above, however in this instance, the choice between causal order can depend upon the history (what has gone on in Alice's lab). Both of these cases are covered by the causal model in definition 3.1.

Figure 3.3: Two possible causal scenarios.

An extension of quantum theory, involving *Process Matrices* [188] which connect laboratories in a non theory specific manner but yet are consistent with standard quantum theory operating inside the laboratories (It is not a requirement that Process Matrices necessarily can be realised in our world) has been shown to violate causal inequalities. Post-selected physics also has this property. In the next chapter we will consider associations between these two formalism's. In this chapter we show that standard quantum theory, while including phenomena such as superpositions of time orderings and causal order, cannot violate a causal inequality. For any quantum experiment where systems are passed between labs, we could therefore replicate the results by a classical causal model of the type defined above.

An alternative approach attempt to get to our results is to start with the full set of process matrices and to investigate which are realisable in quantum mechanics. One suggestion is to add a purification postulate [200], similar to [201], and only consider a process physically realisable if it has such a purification. However, this still allows violations of causal inequalities for three parties, and, the details of physically realisable is dependent on the particular purification. It has also been found that some causally non-separable processes admit a causal model [202]. Note that our results do not imply that quantum process matrices must be separable. We instead recover the desired result by only appealing to text book quantum mechanics and causal reasoning.

### 3.3 Quantum Process

We will specify the framework in more detail in this section. The idea is to consider the correlations between  $N$  parties who will perform measurements and transformations of a quantum system. We not only allow the parties to make quantum measurements, and have access to arbitrary ancilla qubits, but we also allow the parties order with which they observe the state to be controlled coherently by a quantum control system. We also permit the systems involved to undergo arbitrary unitary evolution in the time between the laboratories, we even allow this evolution to be adaptive in the sense that it can change depending on the causal order and measurements that have already been made at any point in the experiment.

To allow the maximum possible interference, and avoid any 'collapse' of the wavefunction (and so prevent interference between the causal orders) we model each measurement the parties make with a unitary interaction between the system and results register that each party has. This is followed by projective measurements upon the results ancilla. This picture of measurements was developed by Von-Neumann [203], and has been shown to be the origin of generalised quantum measurements (POVM's) arising from the more primitive projective measurements that appear in postulate four in the introduction. In doing this, we place ourselves in the scenario where the parties can be imagined to come together at the end of the experiment and read the results from their measurement device much, much later than the measurement interaction took place.

Each party in the experiment also has a 'flag' system. This is an additional quantum system,

which keeps track of how many times they have interacted with the measured system. At the end of the experiment, we require that each of the  $N$  parties has interacted with the system exactly once.

The Hilbert space of this experiment can be decomposed into the following elements:

- An arbitrary quantum system  $\mathcal{H}_S$ , which is passed between parties.
- A quantum control  $\mathcal{H}_C$ , which has dimension  $N + 1$ . The basis of this control  $|1\rangle, \dots, |N\rangle$  denote which party will measure next. The state  $|0\rangle$  is reserved for a ‘do nothing’ command. By considering superpositions of this control and how they evolve, we can set up the superpositions of causal order.
- A result register  $\mathcal{H}_{R_i}$  for each of the  $N$  parties. The different results are represented by an orthonormal basis  $|a_i\rangle$ , with  $a_i \in \mathcal{A}_i$  - the alphabet of measurement outcomes, leading to  $\mathcal{H}_{R_i}$  having dimension  $|\mathcal{A}_i| + 1$ . Initially, the results registers are set to the  $|0\rangle_{R_i}$  state, signifying no measurement has taken place yet.
- A flag  $\mathcal{H}_{F_i}$  for each of the  $N$  parties. This will indicate if the party has measured yet or not, and how many times they have interacted with the system. For simplicity, and to make the evolution operator simple to describe as a unitary, we make these systems infinite dimensional. Basis states will be labelled by the integers. Then the party interacts with the system, the value of the flag is raised by the operator  $\Gamma = \sum_n |n+1\rangle_{F_i} \langle n|_{F_i}$ . Each flag starts in the  $|0\rangle_{F_i}$  state, and at the end of the protocol all flags must be in the  $|1\rangle_{F_i}$  state.

Note that any ancillas that can be used as a working space by the parties can be considered a part of  $\mathcal{H}_s$ . As later parties have all the history of a previous party anyway, this will not be a restriction that we want to impose. We will denote the combined results and flag spaces by  $\mathcal{H}_R = \otimes \mathcal{H}_{R_i}$  and  $\mathcal{H}_F = \otimes \mathcal{H}_{F_i}$  respectively.

Our general quantum protocol takes the following form. We start by initialising the state

$$(3.4) \quad |0\rangle = |0\rangle_S |0\rangle_C |0\rangle_R |0\rangle_F \in \mathcal{H}_S \otimes \mathcal{H}_C \otimes \mathcal{H}_R \otimes \mathcal{H}_F$$

and each party  $l$  is independently issued their classical measurement setting information  $x_l$ . The protocol then proceeds over  $T$  time-steps labelled by an integer  $t$ , each of which is composed of two operations. Firstly, an arbitrary unitary transformation  $\hat{U}_t$  is applied to the the system and control. This unitary can depend on the time-step  $t$ . Next, a fixed controlled lab activation unitary  $\hat{V}$  is applied, which activates whatever party is specified by the state of the control to perform their measurement and transformation sub-routine coherently. This unitary is given by

$$(3.5) \quad \hat{V} = |0\rangle \langle 0|_C \otimes \mathbb{1} + \sum_{l=1}^N |l\rangle \langle l|_C \otimes \hat{V}_{S,R_l}(x_l) \otimes \Gamma_{F_l} \otimes \mathbb{1}$$

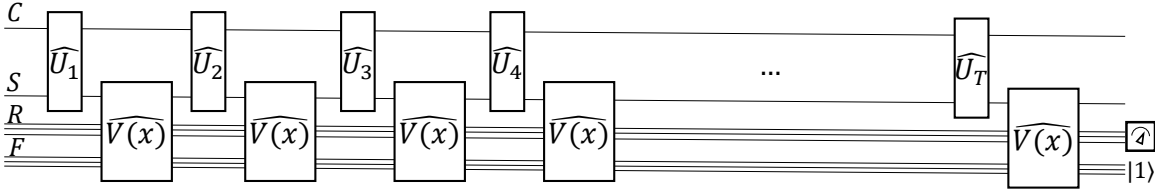


Figure 3.4: Quantum circuit description of the generic measurement procedure for quantum theory. On the left-hand side, we have the initialisation step, where the Control, System, Results and Flag registers are initialised to the zero state. A sequence of  $\hat{U}$  and  $\hat{V}$  operators are then applied in order to measure the system. The measurements are achieved by an entangling unitary between the system, results and flag spaces for each party. At the end of the protocol, we demand that the flags are all raised, and then a projective measurement is performed on the results registers. By repeating this procedure many times,  $p^{\text{quantum}}(\vec{a}|\vec{x})$  are constructed.

Where the identities are implicit over all remaining subsystems.  $\hat{V}_{S,R_l}(x_l)$  is a unitary which implements the measurement of party  $l$  on the system, and is a function of the classical measurement setting data  $x_l$ . This measurement unitary is entangling, and stores the result of the measurement in the register  $R_l$ . The operator  $\Gamma_{F_l}$  increments the flag system of the party making their measurement, when they do. Note that none of the parties have access to an operation which resets the flag, aside from the initialisation operation at the start of the protocol. They therefore always ‘remember’ if they have made a measurement or not. At the end of the protocol, we require that the flags are in the state  $|1\rangle_F^N$  (i.e. that each party has measured the system once, and is sure of this fact). This places constraints on the possible protocols which can be constructed. This restriction is a constraint on the sequence of allowed unitaries  $U_t$ . The total unitary for the protocol is given by

$$(3.6) \quad \mathcal{U} = VU_T V \dots VU_t V \dots VU_1.$$

At the end of the protocol, each party performs a projective measurement on their results register to obtain a final result. The output probability distribution of the quantum protocol is therefore given by

$$(3.7) \quad p^{\text{quantum}}(\vec{a}|\vec{x}) = |(\langle \vec{a} | \langle \vec{a} |_r \otimes I) \mathcal{U} | 0 \rangle|^2.$$

### 3.4 Main Result

The main result in this chapter can be surmised in the following theorem.

**Theorem 3.1.** *Any quantum probability distribution  $p^{\text{quantum}}(\vec{a}|\vec{x})$  can be exactly replicated by a classically causal process  $p^{\text{causal}}(\vec{a}|\vec{x})$ . Hence quantum theory cannot violate a causal inequality.*

The theorem states that any probability distribution generated by the protocol outlined in the previous section, could have also been obtained by a classically causal process (one which

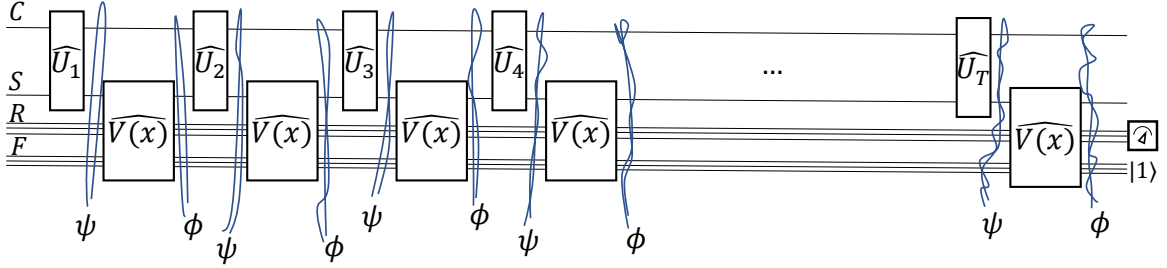


Figure 3.5: A schematic of where the states  $\phi$  and  $\psi$  appearing in definitions 3.3 and 3.2 live in the protocol.

permits no superposition between causal orders). In particular, we will show how to construct an explicit causal process which replicates the probabilities of any quantum protocol perfectly. We will begin with a few definitions, setting up the proof. First, we define a notation for describing states of the entire experiment at each stage of the protocol.

**Definition 3.2.** *The (un-normalised) state with a History  $H_{k-1}$ , at a time  $t$ , with the control set to trigger the action of party  $l_k$  is given by*

$$(3.8) \quad |\psi_{(l_k, t, H_{k-1})}\rangle = (|l_k\rangle \langle l_k|_C \otimes \pi_{RF}^{H_{k-1}} \otimes \mathbb{1}_S) U_t V U_{t-1} \dots V U_1 |0\rangle.$$

The projector onto the result and flag spaces is given by

$$(3.9) \quad \pi_{RF}^{H_{k-1}} = \bigotimes_{i=1}^N \left( \pi_{R_i F_i}^{H_{k-1}} \right),$$

where

$$(3.10) \quad \pi_{R_i F_i}^{H_{k-1}} = \begin{cases} |a_i\rangle \langle a_i|_{R_i} \otimes |1\rangle \langle 1|_{F_i} & \text{if } (i, a_i, x_i) \in H_{k-1}, \\ I_{R_i} \otimes |0\rangle \langle 0|_{F_i} & \text{otherwise.} \end{cases}$$

This defines states which are about to be measured. In other words, a state which is about to undergo a  $\hat{V}$  type operation. The corresponding states following the action of a  $\hat{V}$  are given as,

**Definition 3.3.** *The (un-normalised) state with a History  $H_k$ , at a time  $t$ , in which party  $l_k$  has just acted is given by*

$$(3.11) \quad |\phi_{(l_k, t, H_k)}\rangle = (|a_{l_k}\rangle \langle a_{l_k}|_{R_k} \otimes \mathbb{1}) V |\psi_{(l_k, t, H_{k-1})}\rangle.$$

The states appearing in definitions 3.3 and 3.2 will be accompanied with the corresponding end of protocol states. These will be useful in the proof as a shorthand notation for the time evolution of these states.

**Definition 3.4.** *The states at the end of the protocol for definition 3.2 are given by*

$$(3.12) \quad |\bar{\psi}_{(l_k, t, H_{k-1})}\rangle = V U_T V U_{T-1} \dots U_{t+1} V |\psi_{(l_k, t, H_{k-1})}\rangle.$$

It will also be convenient for the proof to define  $|\psi_{(0,t,H_{k-1})}\rangle$  and  $|\bar{\psi}_{(0,t,H_{k-1})}\rangle$ , which are the same as the above states, but with  $l_k = 0$  (i.e. the control in the ‘do nothing’ setting).

And also for the states after measurement

**Definition 3.5.** We also define the state after measurement, evolved to the end of protocol to be

$$(3.13) \quad |\bar{\phi}_{(l_k,t,H_k)}\rangle = VU_T VU_{T-1} \dots U_{t+1} |\phi_{(l_k,t,H_k)}\rangle.$$

It will also be convenient for the proof to define  $|\phi_{(0,t,H_k)}\rangle = V |\psi_{(0,t,H_k)}\rangle$  and

$$|\bar{\phi}_{(0,t,H_k)}\rangle = VU_T VU_{T-1} \dots U_{t+1} |\phi_{(0,t,H_k)}\rangle.$$

By these definitions, we can associate the states to probabilities arising in a causal model (3.1).

**Definition 3.6.** The probability for party  $l_k$  to act next, given a history  $H_{k-1}$  is given by:

$$(3.14) \quad p_k(l_k | H_{k-1}) = \frac{\sum_{t_k=1}^T ||\psi_{(l_k,t_k,H_{k-1})}\rangle|^2}{\sum_{l'_k \notin \mathcal{L}_{k-1}} \sum_{t'_k=1}^T ||\psi_{(l'_k,t'_k,H_{k-1})}\rangle|^2}.$$

We should take a moment to pause and examine this association. This probability represents the chance for party  $l_k$  to be at stage  $k$  of the casual order, conditioned on a history  $H_{k-1}$ . In the numerator, we have taken the modulus squared of all states where the history  $H_{k-1}$  has occurred and the control is set to trigger party  $l_k$  in the next stage of the protocol. We have also summed over time-steps. Doing this respects the fact that within the quantum paradigm we allow the same measurement sequence to occur at different time-steps. At no point do we provide access to the parties any background clock, and so it is possible measurements can be made to occur coherently over many times. Note that states at different times combine incoherently, but different sequences leading to the same set of historical measurement results combine coherently inside  $|\psi_{(l_k,t_k,H_{k-1})}\rangle$ .

Equation (3.14) also manifestly defines a probability distribution. All terms are non-negative and hence so is the overall summation, and also it is correctly normalised so that  $\sum_{l_k \notin \mathcal{L}_{k-1}} p_k(l_k | H_{k-1}) = 1$ . Also note that it depends on only those input variables  $x_i$  which appear in the history  $H_{k-1}$ . Next, we specify similar probabilities for seeing measurement results based on a given history.

**Definition 3.7.** The probability for party  $l_k$  to obtain the measurement result  $a_{l_k}$ , given a history  $H_{k-1}$ , and an input variable  $x_{l_k}$  is given by:

$$(3.15) \quad p_k(a_{l_k} | H_{k-1}, x_{l_k}) = \frac{\sum_{t_k=1}^T ||\phi_{(l_k,t_k,H_k)}\rangle|^2}{\sum_{a'_{l_k} \in \mathcal{A}_{l_k}} \sum_{t'_k=1}^T ||\phi_{(l_k,t'_k,H'_k)}\rangle|^2},$$

where  $H_k = (H_{k-1}, (l_k, a_{l_k}, x_{l_k}))$  and  $H'_k = (H_{k-1}, (l_k, a'_{l_k}, x_{l_k}))$



Again, (3.15) manifestly are probabilities with the correct normalisation obeyed,  $\sum_{a_{l_k} \in \mathcal{A}_{l_k}} p_k(a_{l_k} | H_{k-1}, x_{l_k}) = 1$ . In the numerator, we have taken a sum of the modulus squared of all of the states which have the correct historical results, irrespective of the background time they were obtained at, the control is in the correct state, and the results register containing the result we want to calculate the probability for, is filled.

Before proving the theorem (3.1) in full generality, it will be helpful to give a brief sketch of the proof. We want to show that the causal model specified by the  $p_k(l_k | H_{k-1})$  and  $p_k(a_{l_k} | H_{k-1}, x_{l_k})$  which are defined above yields the correct quantum probability distribution  $p^{\text{quantum}}(\vec{a} | \vec{x})$  given by (3.7).

We begin by inserting  $p_k(l_k | H_{k-1})$  (from (3.14)) and  $p_k(a_{l_k} | H_{k-1}, x_{l_k})$  (from (3.15)) into the definition of a causal model (3.1). We are then able to straightforwardly cancel the numerator of the ‘who is next?’ type probabilities with the denominator of the ‘what did they see?’ terms for the probabilities evaluated *at the same stage of the causal order*,

$$(3.16) \quad p_k(l_k | H_{k-1}) p_k(a_{l_k} | x_{l_k}, H_{k-1}) = \frac{\sum_{t_k=1}^T |\langle \phi_{(l_k, t_k, H_k)} \rangle|^2}{\sum_{l'_k \notin \mathcal{L}_{k-1}} \sum_{t'_k=1}^T |\langle \psi_{(l'_k, t'_k, H_{k-1})} \rangle|^2}.$$

This allows us to write terms for each stage of the causal order in a compact form. We will then employ a technical result, which allows us to cancel the numerator of  $p_k(l_k | H_{k-1}) p_k(a_{l_k} | x_{l_k}, H_{k-1})$  with the denominator of  $p_{k+1}(l_{k+1} | H_k) p_{k+1}(a_{l_{k+1}} | x_{l_{k+1}}, H_k)$ , so that we can also cancel terms *between different stages of the causal order*. Next, we make the observation that for the first stage of the causal order, the denominator of  $p_1(l_1 | H_0)$  satisfies  $\sum_{l_1 \notin \mathcal{L}_0} \sum_{t_1=1}^T |\langle \psi_{(l_1, t_1, H_0)} \rangle|^2 = 1$  (which corresponds to the fact that someone must measure first in the quantum circuit). Finally, we note that the numerator of  $p_N(l_N | H_{N-1}) p_N(a_{l_N} | x_{l_N}, H_{N-1})$ , summed over  $l_N$ , represents exactly the probabilities  $p^{\text{quantum}}(\vec{a} | \vec{x})$  arising from the quantum protocol. This allows us to simulate the results of the quantum protocol via the classically causal model given in (3.1).

The final definition we will be needing is a formal definition of the protocol. This follows from the discussion in the previous chapter.

**Definition 3.8.** *The quantum protocol consists of preparing an initial state  $|0\rangle$ , then acting with an alternating sequence of unitaries  $U_t$  that act on the system and the control, and unitaries  $V$  that act on the system, results and flag spaces as specified by the control. The total unitary for the protocol is given by*

$$(3.17) \quad \mathcal{U} = VU_T VU_{T-1} V \dots VU_N V \dots VU_1$$

where we note that for an  $N$  party protocol,  $T \geq N$ . Finally, the results registers are measured in a basis, giving the outcome probability distribution

$$(3.18) \quad p^{\text{quantum}}(\vec{a} | \vec{x}) = |(\vec{a}) \langle \vec{a} |_R \otimes \mathbb{1} \rangle \mathcal{U} |0\rangle|^2.$$

With these definitions in place, we first prove some useful orthogonality lemmas concerning the barred states. We will make use of these extensively throughout the following proof.

**Lemma 3.1.** *We have that*

$$(3.19) \quad \langle \bar{\psi}_{l',t',H} | \bar{\psi}_{l,t,H} \rangle = 0$$

unless  $l = l'$  and  $t' = t$ .

Consider first that  $t = t'$  and  $l \neq l'$ . Then we have that  $\langle \bar{\psi}_{l',t,H} | \bar{\psi}_{l,t,H} \rangle = \langle \psi_{l',t,H} | \psi_{l,t,H} \rangle = 0$ , since  $|\psi_{l',t,H}\rangle$  and  $|\psi_{l,t,H}\rangle$  are orthogonal on the control  $\mathcal{H}_c$ . Next, consider that  $t < t'$ . Then  $\langle \bar{\psi}_{l',t',H} | \bar{\psi}_{l,t,H} \rangle = \langle \psi_{l',t',H} | U_{t'} V \dots U_{t+1} V | \psi_{l,t,H} \rangle = 0$  since  $V | \psi_{l,t,H} \rangle$  contains a raised  $l$  flag that is not raised in  $\langle \psi_{l',t',H} |$ , and there is no operator connecting the two which can lower this flag. The case with  $t > t'$  follows from the  $t < t'$  case by noting that  $\langle \bar{\psi}_{l',t',H} | \bar{\psi}_{l,t,H} \rangle = \langle \bar{\psi}_{l,t,H} | \bar{\psi}_{l',t',H} \rangle^*$ .

**Lemma 3.2.** *We have that*

$$(3.20) \quad \langle \bar{\phi}_{l',t',H} | \bar{\phi}_{l,t,H} \rangle = 0$$

unless  $l = l'$  and  $t = t'$ .

Consider first that  $t = t'$  and  $l \neq l'$ . Then we have that  $\langle \bar{\phi}_{l',t,H} | \bar{\phi}_{l,t,H} \rangle = \langle \phi_{l',t,H} | \phi_{l,t,H} \rangle = 0$ , since  $|\phi_{l',t,H}\rangle$  and  $|\phi_{l,t,H}\rangle$  are orthogonal on the control  $\mathcal{H}_c$ . Next, consider that  $t < t'$ . Then  $\langle \bar{\phi}_{l',t',H} | \bar{\phi}_{l,t,H} \rangle = \langle \phi_{l',t',H} | V U_{t'} V \dots U_{t+1} | \phi_{l,t,H} \rangle = 0$ , since the leftmost  $V$  either raises a flag not in the history  $H$ , or the control at this point is set to zero, either of which will give the desired orthogonality. The case with  $t > t'$  follows from the  $t < t'$  case by noting that  $\langle \bar{\phi}_{l',t',H} | \bar{\phi}_{l,t,H} \rangle = \langle \bar{\phi}_{l,t,H} | \bar{\phi}_{l',t',H} \rangle^*$ .

We now move onto proving the main result in full generality. This proof has four stages,

1. The first concerns a cancellation within terms of the same causal order stage, which allows us to rewrite the causal model in a nice way.
2. The second result concerns the initial term in the inductive proof. This corresponds to a mathematical paraphrasing ‘somebody has to measure first’ in the quantum protocol.
3. The third result concerns final terms in the inductive proof. We will show that the final terms in the causal model have sufficient expressive power to obtain the quantum measurement probabilities in their entirety.
4. The fourth result concerns cancellations between terms at subsequent stages of the causal order. This leads to our main result which ties all of this together for a full proof that  $p(\vec{a}|\vec{x}) = |\langle \vec{a} | \mathcal{U} | 0 \rangle|^2$  is causal.

**Result 3.1.** *There is an equality between the numerator of the ‘who is next?’ type probabilities  $p_k(l_k | H_{k-1})$ , and the denominator of the ‘results’ type probabilities  $p_k(a_{l_k} | H_{k-1}, x_{l_k})$ , allowing us to write the product of these probabilities in a nice way*

$$(3.21) \quad p_k(l_k | H_{k-1}) p_k(a_{l_k} | x_{l_k}, H_{k-1}) = \frac{\sum_{t_k=1}^T |\langle \phi_{(l_k, t_k, H_k)} \rangle|^2}{\sum_{l'_k \in \mathcal{L}_{k-1}} \sum_{t'_k=1}^T |\langle \psi_{(l'_k, t'_k, H_{k-1})} \rangle|^2}.$$

Proof: starting with the denominator of the ‘results’ probability

$$\begin{aligned}
 \sum_{a'_{l_k} \in \mathcal{A}_{l_k}} \sum_{t_k=1}^T \left| |\phi_{(l_k, t_k, H'_k)}\rangle \right|^2 &= \sum_{a'_{l_k} \in \mathcal{A}_{l_k}} \sum_{t_k=1}^T \left| (|a'_{l_k}\rangle \langle a'_{l_k}|_{r_k} \otimes \mathbb{1}) V |\psi_{(l_k, t_k, H_{k-1})}\rangle \right|^2 \\
 &= \sum_{t_k=1}^T \left| \sum_{a'_{l_k} \in \mathcal{A}_{l_k}} (|a'_{l_k}\rangle \langle a'_{l_k}|_{r_k} \otimes \mathbb{1}) V |\psi_{(l_k, t_k, H_{k-1})}\rangle \right|^2. \\
 &= \sum_{t_k=1}^T |V |\psi_{(l_k, t_k, H_{k-1})}\rangle|^2 \\
 (3.22) \qquad \qquad \qquad &= \sum_{t_k=1}^T ||\psi_{(l_k, t_k, H_{k-1})}\rangle|^2,
 \end{aligned}$$

we obtain the numerator of the ‘who is next?’ probabilities. In the second line we have used orthogonality on the results register, in the third line we have used the fact that after a measurement by party  $l_k$ , some result in  $\mathcal{A}_{l_k}$  must have been obtained, and in the final line we have used unitarity. Using this to cancel the numerator of (3.14) with the denominator of (3.15) we obtain the desired result.

**Result 3.2.** *The denominator of the first term  $p_1(l_1|H_0)$  satisfies*

$$(3.23) \qquad \sum_{l_1 \notin \mathcal{L}_0} \sum_{t_1=1}^T ||\psi_{(l_1, t_1, H_0)}\rangle|^2 = 1.$$

Proof: by first using unitarity and then Lemma 3.1 we have

$$\begin{aligned}
 \sum_{l_1 \notin \mathcal{L}_0} \sum_{t_1=1}^T ||\psi_{(l_1, t_1, H_0)}\rangle|^2 &= \sum_{l_1 \notin \mathcal{L}_0} \sum_{t_1=1}^T ||\bar{\psi}_{(l_1, t_1, H_0)}\rangle|^2 \\
 (3.24) \qquad \qquad \qquad &= \left| \sum_{l_1 \notin \mathcal{L}_0} \sum_{t_1=1}^T |\bar{\psi}_{(l_1, t_1, H_0)}\rangle \right|^2.
 \end{aligned}$$

To simplify this further, consider evolving the state  $|\psi_{0, t_1-1, H_0}\rangle$  forward for a full time-step using  $U_{t_1} V$ . As the control is in state 0, no measurement occurs during  $V$ , and the unitary  $U_1$  creates a superposition in which the control takes any possible state. Symbolically,

$$(3.25) \qquad U_{t_1} V |\psi_{0, t_1-1, H_0}\rangle = \sum_{l_1 \notin \mathcal{L}_0} |\psi_{(l_1, t_1, H_0)}\rangle + |\psi_{(0, t_1, H_0)}\rangle.$$

By applying  $V U_T V \dots U_{t_1+1} V$  to this equation, we can obtain a similar relation for the barred states,

$$(3.26) \qquad |\bar{\psi}_{0, t_1-1, H_0}\rangle = \sum_{l_1 \notin \mathcal{L}_0} |\bar{\psi}_{(l_1, t_1, H_0)}\rangle + |\bar{\psi}_{(0, t_1, H_0)}\rangle.$$

We can rearrange this equation and substitute for the sum over  $l_1$  on the right-hand side of (3.24) to obtain

$$(3.27) \qquad \sum_{l_1 \notin \mathcal{L}_0} \sum_{t_1=1}^T ||\psi_{(l_1, t_1, H_0)}\rangle|^2 = \left| \sum_{t_1=1}^T (|\bar{\psi}_{(0, t_1-1, H_0)}\rangle - |\bar{\psi}_{(0, t_1, H_0)}\rangle) \right|^2$$

By expanding the summation on the right-hand side we find that only the first and last terms remain, giving

$$(3.28) \quad \sum_{l_1 \notin \mathcal{L}_0} \sum_{t_1=1}^T |\psi_{(l_1, t_1, H_0)}\rangle|^2 = |\bar{\psi}_{(0,0,H_0)}\rangle - |\bar{\psi}_{(0,T,H_0)}\rangle|^2$$

Note that it is impossible by the requirements of our protocol that no-one has measured by time  $t = T$ . Such a scenario would violate the assumption that there are exactly  $N$  flags raised at the end of the protocol. Therefore,  $|\bar{\psi}_{(0,T,H_0)}\rangle = 0$ , and we find

$$(3.29) \quad \sum_{l_1 \notin \mathcal{L}_0} \sum_{t_1=1}^T |\psi_{(l_1, t_1, H_0)}\rangle|^2 = |\bar{\psi}_{(0,0,H_0)}\rangle|^2 = |\mathcal{U}|0\rangle|^2 = 1$$

as desired.

**Result 3.3.** *The outcome statistics in the numerator of the final term in the causal probabilistic model represent the quantum probabilities arising from the protocol. In other words,*

$$(3.30) \quad \sum_{l_N \in \mathcal{L}_N} \sum_{t_N=1}^T |\phi_{(l_N, t_N, H_N)}\rangle|^2 = |(\vec{a}) \langle \vec{a} | \otimes \mathbb{1} \mathcal{U} |0\rangle|^2$$

*Proof:* Firstly, note that by unitarity and Lemma 3.2 we have that

$$(3.31) \quad \begin{aligned} \sum_{l_N \in \mathcal{L}_N} \sum_{t_N=1}^T |\phi_{(l_N, t_N, H_N)}\rangle|^2 &= \sum_{l_N \in \mathcal{L}_N} \sum_{t_N=1}^T |\bar{\phi}_{(l_N, t_N, H_N)}\rangle|^2 \\ &= \left| \sum_{l_N \in \mathcal{L}_N} \sum_{t_N=1}^T |\bar{\phi}_{(l_N, t_N, H_N)}\rangle \right|^2. \end{aligned}$$

The history  $H_N$  represents a case in which all parties have already measured. At time  $t < T$ , what are the possible ways that this history can be filled? Either, nobody has measured in the previous time-step, or the last party to be filled into the history (subject to the requirement every party must enter the history exactly once) has just measured. In any case, evolving the linear combination of these states forward a time-step must produce a state at  $t + 1$  which contains an empty control (i.e., no-one else is left to measure, so don't trigger them!). This gives us the key relation

$$(3.32) \quad VU_{t+1} \left( \sum_{l_N \in \mathcal{L}_N} |\phi_{(l_N, t, H_N)}\rangle + |\phi_{(0, t, H_N)}\rangle \right) = |\phi_{(0, t+1, H_N)}\rangle$$

which holds for  $t < T$ . Applying  $VU_T V \dots U_{t+2}$  we obtain

$$(3.33) \quad \left( \sum_{l_N \in \mathcal{L}_N} |\bar{\phi}_{(l_N, t, H_N)}\rangle + |\bar{\phi}_{(0, t, H_N)}\rangle \right) = |\bar{\phi}_{(0, t+1, H_N)}\rangle$$

which we can then rearrange to get

$$(3.34) \quad \sum_{l_N \in \mathcal{L}_N} |\bar{\phi}_{(l_N, t, H_N)}\rangle = |\bar{\phi}_{(0, t+1, H_N)}\rangle - |\bar{\phi}_{(0, t, H_N)}\rangle.$$

By separating out the  $t_N = T$  term in equation (3.31) and then substituting this in the remaining terms, we find that

$$\begin{aligned}
 \sum_{l_N \in \mathcal{L}_N} \sum_{t_N=1}^T ||\phi_{(l_N, t_N, H_N)}\rangle|^2 &= | \sum_{l_N \in \mathcal{L}_N} |\bar{\phi}_{(l_N, T, H_N)}\rangle + \sum_{t_N=1}^{T-1} \sum_{l_N \in \mathcal{L}_N} |\bar{\phi}_{(l_N, t_N, H_N)}\rangle|^2 \\
 &= | \sum_{l_N \in \mathcal{L}_N} |\bar{\phi}_{(l_N, T, H_N)}\rangle + \sum_{t_N=1}^{T-1} (|\bar{\phi}_{(0, t_N+1, H_N)}\rangle - |\bar{\phi}_{(0, t_N, H_N)}\rangle) |^2 \\
 (3.35) \qquad \qquad \qquad &= | \sum_{l_N \in \mathcal{L}_N} |\bar{\phi}_{(l_N, T, H_N)}\rangle + |\bar{\phi}_{(0, T, H_N)}\rangle - |\bar{\phi}_{(0, 1, H_N)}\rangle |^2.
 \end{aligned}$$

Now we note that  $|\bar{\phi}_{(0, 1, H_N)}\rangle = 0$  since it would be impossible for all parties to have measured in one time-step, and for the control to be in the zero state. Then we note that  $\sum_{l_N \in \mathcal{L}_N} |\bar{\phi}_{(l_N, T, H_N)}\rangle + |\bar{\phi}_{(0, T, H_N)}\rangle = (\pi_{rf}^{H_N} \otimes \mathbb{1}) \mathcal{U} |0\rangle$ , which is to say that these are simply the possible states at the end of the protocol, containing the measurement results we want to calculate the probabilities for in the history. Therefore

$$\begin{aligned}
 \sum_{l_N \in \mathcal{L}_N} \sum_{t_N=1}^T ||\phi_{(l_N, t_N, H_N)}\rangle|^2 &= |(\pi_{rf}^{H_N} \otimes \mathbb{1}) \mathcal{U} |0\rangle|^2 \\
 (3.36) \qquad \qquad \qquad &= |(|\vec{a}\rangle \langle \vec{a}| \otimes \mathbb{1}) \mathcal{U} |0\rangle|^2
 \end{aligned}$$

as desired.

**Result 3.4.** *This is a technical result which establishes an equality between the states after measurement at causal order stage  $k$  and the states before measurement at the next stage of the causal order. Namely, for  $1 \leq k < N$  that:*

$$(3.37) \qquad \sum_{l_k \in \mathcal{L}_k} \sum_{t=1}^T ||\phi_{(l_k, t, H_k)}\rangle|^2 = \sum_{t'=1}^T \sum_{l'_{k+1} \notin \mathcal{L}_k} ||\psi_{(l'_{k+1}, t', H_k)}\rangle|^2.$$

Proof: Firstly, by unitarity and Lemma 3.2 we have that

$$\begin{aligned}
 \sum_{l_k \in \mathcal{L}_k} \sum_{t=1}^T ||\phi_{(l_k, t, H_k)}\rangle|^2 &= \sum_{l_k \in \mathcal{L}_k} \sum_{t=1}^T ||\bar{\phi}_{(l_k, t, H_k)}\rangle|^2 \\
 (3.38) \qquad \qquad \qquad &= | \sum_{l_k \in \mathcal{L}_k} \sum_{t=1}^T |\bar{\phi}_{(l_k, t, H_k)}\rangle |^2.
 \end{aligned}$$

Consider time-evolving a state just after the  $t^{\text{th}}$  measurement step, in which history  $H_k$  has been obtained (either by the last party just having measured, or by all parties in  $H_k$  having measured previously), by  $U_{t+1}$ . This links states of the form  $|\phi_{(l_k, t, H_k)}\rangle$  and  $|\psi_{(l'_{k+1}, t+1, H_k)}\rangle$  via

$$(3.39) \qquad U_{t+1} \left( \sum_{l_k \in \mathcal{L}_k} |\phi_{(l_k, t, H_k)}\rangle + |\phi_{(0, t, H_k)}\rangle \right) = \sum_{l'_{k+1} \notin \mathcal{L}_k} |\psi_{(l'_{k+1}, t+1, H_k)}\rangle + |\psi_{(0, t+1, H_k)}\rangle$$

for  $t < T$ . Applying  $VU_T V \dots U_{t+2} V$  we obtain a very similar result for barred states;

$$(3.40) \quad \sum_{l_k \in \mathcal{L}_k} |\bar{\phi}(l_k, t, H_k)\rangle + |\bar{\phi}(0, t, H_k)\rangle = \sum_{l'_{k+1} \notin \mathcal{L}_k} |\bar{\psi}(l'_{k+1}, t+1, H_k)\rangle + |\bar{\psi}(0, t+1, H_k)\rangle.$$

We also note that  $|\phi(0, t, H_k)\rangle = V |\psi(0, t, H_k)\rangle$  and hence that  $|\bar{\phi}(0, t, H_k)\rangle = |\bar{\psi}(0, t, H_k)\rangle$ . Making this substitution and rearranging a little we get

$$(3.41) \quad \sum_{l_k \in \mathcal{L}_k} |\bar{\phi}(l_k, t, H_k)\rangle = \sum_{l'_{k+1} \notin \mathcal{L}_k} |\bar{\psi}(l'_{k+1}, t+1, H_k)\rangle + |\bar{\psi}(0, t+1, H_k)\rangle - |\bar{\psi}(0, t, H_k)\rangle.$$

By substituting (3.41) into (3.38) for  $t < T$ , we arrive at

$$(3.42) \quad \begin{aligned} \sum_{l_k \in \mathcal{L}_k} \sum_{t=1}^T \left| |\phi(l_k, t, H_k)\rangle \right|^2 &= \left| \sum_{l_k \in \mathcal{L}_k} |\bar{\phi}(l_k, T, H_k)\rangle + \sum_{t=1}^{T-1} \left( \sum_{l'_{k+1} \notin \mathcal{L}_k} |\bar{\psi}(l'_{k+1}, t+1, H_k)\rangle + |\bar{\psi}(0, t+1, H_k)\rangle - |\bar{\psi}(0, t, H_k)\rangle \right) \right|^2 \\ &= \left| \sum_{l_k \in \mathcal{L}_k} |\bar{\phi}(l_k, T, H_k)\rangle + \sum_{t=1}^{T-1} \sum_{l'_{k+1} \notin \mathcal{L}_k} |\bar{\psi}(l'_{k+1}, t+1, H_k)\rangle + |\bar{\psi}(0, T, H_k)\rangle - |\bar{\psi}(0, 1, H_k)\rangle \right|^2 \end{aligned}$$

where for the sums over time in the last two terms only the states with maximal and minimal times remain.

Given that  $k < N$  and all parties must have measured by the end of the protocol, it must be the case that  $|\bar{\phi}(l_k, T, H_k)\rangle = 0$  and  $|\bar{\psi}(0, T, H_k)\rangle = 0$ . Also as  $k \geq 1$  it must be the case that  $|\bar{\psi}(0, 1, H_k)\rangle = 0$  and  $|\bar{\psi}(l'_{k+1}, 1, H_k)\rangle = 0$ , as these states are just before the first measurement and hence must have no history. Using these results in equation (3.42) and setting  $t' = t + 1$ , we obtain

$$(3.43) \quad \sum_{l_k \in \mathcal{L}_k} \sum_{t=1}^T \left| |\phi(l_k, t, H_k)\rangle \right|^2 = \left| \sum_{t'=1}^T \sum_{l'_{k+1} \notin \mathcal{L}_k} |\bar{\psi}(l'_{k+1}, t', H_k)\rangle \right|^2.$$

Finally, using Lemma 3.1 and unitarity we arrive at

$$(3.44) \quad \begin{aligned} \sum_{l_k \in \mathcal{L}_k} \sum_{t=1}^T \left| |\phi(l_k, t, H_k)\rangle \right|^2 &= \sum_{t'=1}^T \sum_{l'_{k+1} \notin \mathcal{L}_k} \left| |\bar{\psi}(l'_{k+1}, t', H_k)\rangle \right|^2 \\ &= \sum_{t'=1}^T \sum_{l'_{k+1} \notin \mathcal{L}_k} \left| |\psi(l'_{k+1}, t', H_k)\rangle \right|^2 \end{aligned}$$

as required. By pulling together the previous results in stages 1 – 4, we can now complete our proof of the main theorem.

**Result 3.5.** *We will now show that the results of the quantum protocol can be replicated by a causal process. In particular*

$$p(\vec{a}|\vec{x}) = |(|\vec{a}\rangle \langle \vec{a}| \otimes \mathbb{1}) \mathcal{U} |0\rangle|^2 =$$

(3.45)

$$\sum_{l_1 \notin \mathcal{L}_0} \sum_{l_2 \notin \mathcal{L}_1} \dots \sum_{l_N \notin \mathcal{L}_{N-1}} p_1(l_1|H_0) p_1(a_{l_1}|H_0, x_{l_1}) p_2(l_2|H_1) p_2(a_{l_2}|H_1, x_{l_2}) \dots p_N(l_N|H_{N-1}) p_N(a_{l_N}|H_{N-1}, x_{l_N})$$

and as such, the outcome statistics  $p(\vec{a}|\vec{x})$  cannot violate a causal inequality.

Proof: Firstly, substituting definitions 3.6 and 3.7 into the causal model (3.1), and then using result 3.1, we can re-write the probability distribution for the entire causal model as

$$(3.46) \quad p(\vec{a}|\vec{x}) = \sum_{l_1 \notin \mathcal{L}_0} \sum_{l_2 \notin \mathcal{L}_1} \dots \sum_{l_N \notin \mathcal{L}_{N-1}} \frac{\sum_{t_1=1}^T |\langle \phi_{(l_1, t_1, H_1)} \rangle|^2}{\sum_{l'_1 \notin \mathcal{L}_0} \sum_{t'_1=1}^T |\langle \psi_{(l'_1, t'_1, H_0)} \rangle|^2} \frac{\sum_{t_2=1}^T |\langle \phi_{(l_2, t_2, H_2)} \rangle|^2}{\sum_{l'_2 \notin \mathcal{L}_1} \sum_{t'_2=1}^T |\langle \psi_{(l'_2, t'_2, H_1)} \rangle|^2} \dots$$

$$\dots \frac{\sum_{t_N=1}^T |\langle \phi_{(l_N, t_N, H_N)} \rangle|^2}{\sum_{l'_N \notin \mathcal{L}_{N-1}} \sum_{t'_N=1}^T |\langle \psi_{(l'_N, t'_N, H_{N-1})} \rangle|^2}$$

Let us begin by performing a simple reshuffling of (3.46)'s numerators and denominators, by writing the denominator of the term associated to causal order stage  $k$  as the denominator of the term associated to  $k-1$ .

$$(3.47) \quad p(\vec{a}|\vec{x}) = \frac{1}{\sum_{l'_1 \notin \mathcal{L}_0} \sum_{t'_1=1}^T |\langle \psi_{(l'_1, t'_1, H_0)} \rangle|^2} \times$$

$$\times \sum_{l_1 \notin \mathcal{L}_0} \sum_{l_2 \notin \mathcal{L}_1} \dots \sum_{l_N \notin \mathcal{L}_{N-1}} \frac{\sum_{t_1=1}^T |\langle \phi_{(l_1, t_1, H_1)} \rangle|^2}{\sum_{l'_2 \notin \mathcal{L}_1} \sum_{t'_2=1}^T |\langle \psi_{(l'_2, t'_2, H_1)} \rangle|^2} \frac{\sum_{t_2=1}^T |\langle \phi_{(l_2, t_2, H_2)} \rangle|^2}{\sum_{l'_3 \notin \mathcal{L}_2} \sum_{t'_3=1}^T |\langle \psi_{(l'_3, t'_3, H_2)} \rangle|^2} \dots$$

$$\dots \frac{\sum_{t_{k-1}=1}^T |\langle \phi_{(l_{k-1}, t_{k-1}, H_{k-1})} \rangle|^2}{\sum_{l'_k \notin \mathcal{L}_{k-1}} \sum_{t'_k=1}^T |\langle \psi_{(l'_k, t'_k, H_{k-1})} \rangle|^2} \dots \sum_{t_N=1}^T |\langle \phi_{(l_N, t_N, H_N)} \rangle|^2$$

by using result (3.2) we have

$$(3.48) \quad p(\vec{a}|\vec{x}) = \sum_{l_1 \notin \mathcal{L}_0} \sum_{l_2 \notin \mathcal{L}_1} \dots \sum_{l_N \notin \mathcal{L}_{N-1}} \frac{\sum_{t_1=1}^T |\langle \phi_{(l_1, t_1, H_1)} \rangle|^2}{\sum_{l'_2 \notin \mathcal{L}_1} \sum_{t'_2=1}^T |\langle \psi_{(l'_2, t'_2, H_1)} \rangle|^2} \frac{\sum_{t_2=1}^T |\langle \phi_{(l_2, t_2, H_2)} \rangle|^2}{\sum_{l'_3 \notin \mathcal{L}_2} \sum_{t'_3=1}^T |\langle \psi_{(l'_3, t'_3, H_2)} \rangle|^2} \dots \sum_{t_N=1}^T |\langle \phi_{(l_N, t_N, H_N)} \rangle|^2$$

now note that we can rewrite the leftmost sum as

$$(3.49) \quad \sum_{l_1 \notin \mathcal{L}_0} = \sum_{\mathcal{L}_1} \sum_{l_1 \in \mathcal{L}_1},$$

where the sum over  $\mathcal{L}_1$  is over all singleton sets  $\{l_1\}$  (and hence has  $N$  terms), and the subsequent sum over  $l_1 \in \mathcal{L}_1$  contains just a single term.

We can use this to rewrite the probability distribution as

$$(3.50) \quad p(\vec{a}|\vec{x}) = \sum_{\mathcal{L}_1} \frac{\sum_{l_1 \in \mathcal{L}_1} \sum_{t_1=1}^T |\langle \phi_{(l_1, t_1, H_1)} \rangle|^2}{\sum_{l'_2 \notin \mathcal{L}_1} \sum_{t'_2=1}^T |\langle \psi_{(l'_2, t'_2, H_1)} \rangle|^2} \sum_{l_2 \notin \mathcal{L}_1} \dots \sum_{l_N \notin \mathcal{L}_{N-1}} \frac{\sum_{t_2=1}^T |\langle \phi_{(l_2, t_2, H_2)} \rangle|^2}{\sum_{l'_3 \notin \mathcal{L}_2} \sum_{t'_3=1}^T |\langle \psi_{(l'_3, t'_3, H_2)} \rangle|^2} \dots \sum_{t_N=1}^T |\langle \phi_{(l_N, t_N, H_N)} \rangle|^2$$

where we have used the fact that the first numerator and denominator do not depend on  $\{l_2, \dots, l_N\}$ .

By application of result (3.4) this is just

$$(3.51) \quad p(\vec{a}|\vec{x}) = \sum_{\mathcal{L}_1} \sum_{l_2 \notin \mathcal{L}_1} \dots \sum_{l_N \notin \mathcal{L}_{N-1}} \frac{\sum_{t_2=1}^T |\langle \phi_{(l_2, t_2, H_2)} \rangle|^2}{\sum_{l'_3 \notin \mathcal{L}_2} \sum_{t'_3=1}^T |\langle \psi_{(l'_3, t'_3, H_2)} \rangle|^2} \dots \sum_{t_N=1}^T |\langle \phi_{(l_N, t_N, H_N)} \rangle|^2$$

we can iterate the same process again using

$$(3.52) \quad \sum_{\mathcal{L}_1} \sum_{l_2 \notin \mathcal{L}_1} = \sum_{\mathcal{L}_2} \sum_{l_2 \in \mathcal{L}_2} .$$

The left-hand side corresponds to first picking  $l_1$  (with  $N$  possibilities) and then picking a different  $l_2$  ( $N - 1$  possibilities), whereas the right-hand side corresponds to first picking a pair of distinct labs  $\mathcal{L}_2$  (with  $N(N - 1)/2$  possibilities) and then picking which of them was last (2 possibilities). This gives

$$(3.53) \quad p(\vec{a}|\vec{x}) = \sum_{\mathcal{L}_2} \sum_{l_2 \in \mathcal{L}_2} \dots \sum_{l_N \notin \mathcal{L}_{N-1}} \frac{\sum_{t_2=1}^T |\langle \phi_{(l_2, t_2, H_2)} \rangle|^2}{\sum_{l'_3 \notin \mathcal{L}_2} \sum_{t'_3=1}^T |\langle \psi_{(l'_3, t'_3, H_2)} \rangle|^2} \dots \sum_{t_N=1}^T |\langle \phi_{(l_N, t_N, H_N)} \rangle|^2$$

which is just

$$(3.54) \quad p(\vec{a}|\vec{x}) = \sum_{\mathcal{L}_2} \frac{\sum_{l_2 \in \mathcal{L}_2} \sum_{t_2=1}^T |\langle \phi_{(l_2, t_2, H_2)} \rangle|^2}{\sum_{l'_3 \notin \mathcal{L}_2} \sum_{t'_3=1}^T |\langle \psi_{(l'_3, t'_3, H_2)} \rangle|^2} \sum_{l_3 \notin \mathcal{L}_2} \dots \sum_{l_N \notin \mathcal{L}_{N-1}} \frac{\sum_{t_3=1}^T |\langle \phi_{(l_3, t_3, H_3)} \rangle|^2}{\sum_{l'_4 \notin \mathcal{L}_3} \sum_{t'_4=1}^T |\langle \psi_{(l'_4, t'_4, H_3)} \rangle|^2} \dots \sum_{t_N=1}^T |\langle \phi_{(l_N, t_N, H_N)} \rangle|^2$$

application of result (3.4) leads to another cancellation, so that we may write now

$$(3.55) \quad p(\vec{a}|\vec{x}) = \sum_{\mathcal{L}_2} \sum_{l_3 \notin \mathcal{L}_2} \dots \sum_{l_N \notin \mathcal{L}_{N-1}} \frac{\sum_{t_3=1}^T |\langle \phi_{(l_3, t_3, H_3)} \rangle|^2}{\sum_{l'_4 \notin \mathcal{L}_3} \sum_{t'_4=1}^T |\langle \psi_{(l'_4, t'_4, H_3)} \rangle|^2} \dots \sum_{t_N=1}^T |\langle \phi_{(l_N, t_N, H_N)} \rangle|^2$$

We can then iterate this process by applying the general result that

$$(3.56) \quad \sum_{\mathcal{L}_k} \sum_{l_{k+1} \notin \mathcal{L}_k} = \sum_{\mathcal{L}_{k+1}} \sum_{l_{k+1} \in \mathcal{L}_{k+1}} ,$$

and cancelling one of the numerators and denominators using result (3.4) until we are left with the final term,

$$(3.57) \quad p(\vec{a}|\vec{x}) = \sum_{\mathcal{L}_N} \sum_{l_N \in \mathcal{L}_N} \sum_{t_N=1}^T |\langle \phi_{(l_N, t_N, H_N)} \rangle|^2$$

the summation  $\sum_{\mathcal{L}_N} = 1$ , as the only term corresponds to  $\mathcal{L}_N = \{1, 2, \dots, N\}$ . Application of result (3.3) then shows that this causal model indeed reproduces the quantum probabilities, i.e. that

$$(3.58) \quad p(\vec{a}|\vec{x}) = |(\vec{a} | \langle \vec{a} | \otimes \mathbb{1}) \mathcal{U} | 0 \rangle|^2$$

as desired. This completes our proof of theorem (3.1).

### 3.5 Quantum Switch

We now will give an example of how our result translates in practice. This will aid clarity, provide a stimulating new insight into an already well studied subject and hopefully show how although



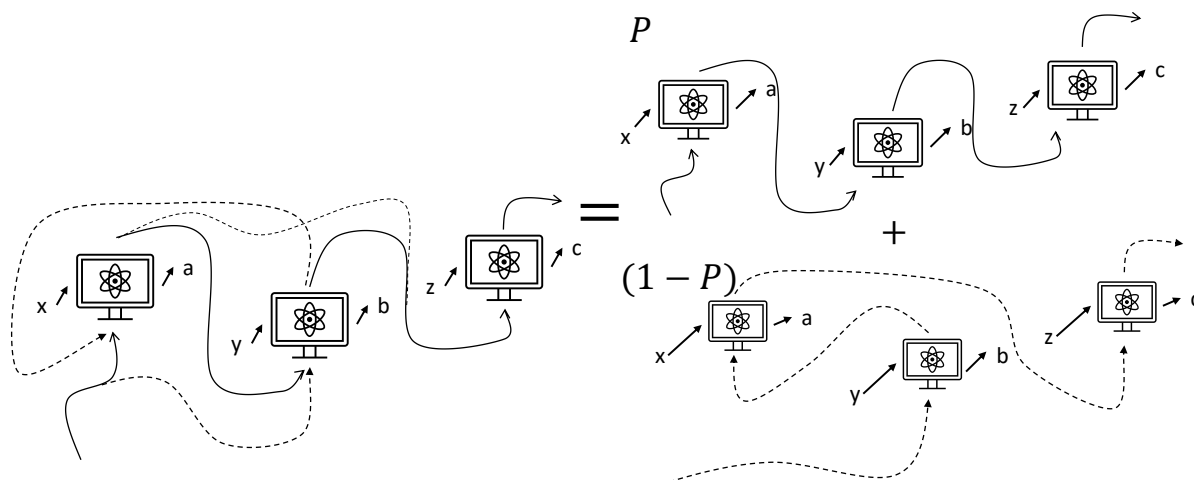


Figure 3.6: Cartoon of what the result says; on the left hand side we have a causally indefinite process, on the right hand side, we have its simulation with a probabilistic causal model.

causal indefiniteness is certainly present within quantum theory, there is no theory independent test for this fact. We will base our example on the quantum switch [179]. The switch uses the state of a quantum control to determine the order in which operations are applied to another quantum system. Here, the operations will be measurements. The switch can be described using a variety of formalism's, for example, as a causally non-separable process matrix [191]. Here, the basic idea will be to prepare a state over the parties of the form

$$(3.59) \quad \frac{1}{\sqrt{2}}(|1\rangle_C \otimes U_A U_B |0\rangle_{SFR} + |2\rangle_C \otimes U_B U_A |0\rangle_{SFR})$$

where  $U_A$  and  $U_B$  are unitary transformations performed by Alice and Bob (representing in our picture measurements). A third party, Charlie, will be responsible for measurements on the control in a basis consisting of superpositions of  $|1\rangle$  and  $|2\rangle$  this will introduce interference between the two causal orders in which either Alice or Bob goes first. It has been shown prior to this work that the quantum switch cannot be used to violate a causal inequality [191, 188] however it is illustrative to see how this result follows from the framework outlined. If more general quantum set-ups than the switch can violate causal inequalities was raised as an open question in [191]. This chapter directly answers in the negative.

Because parties cannot directly access the control system during our protocol, we transfer the state of the control into the system before Charlie's measurement, splitting the system into qubits to facilitate this. We consider the circuit in figure 3.7 as our implementation of the switch, with

$$(3.60) \quad \begin{aligned} U_1 |0\rangle_C &= \frac{1}{\sqrt{2}}(|1\rangle_C + |2\rangle_C), & U_2 |1\rangle_C &= |2\rangle_C, & U_2 |2\rangle_C &= |1\rangle_C, \\ U_3 |1\rangle_C |\psi\rangle_{S_1} |0\rangle_{S_2} &= |3\rangle_C |0\rangle_{S_1} |\psi\rangle_{S_2}, & U_3 |2\rangle_C |\psi\rangle_{S_1} |0\rangle_{S_2} &= |3\rangle_C |1\rangle_{S_1} |\psi\rangle_{S_2}, \end{aligned}$$

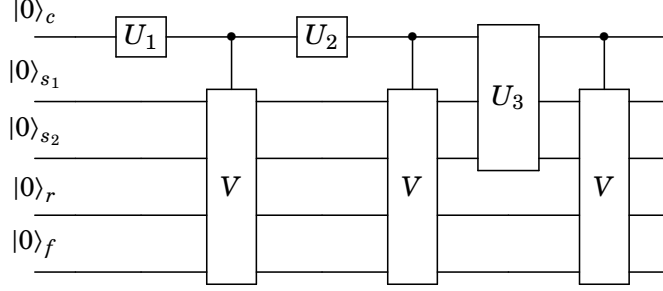


Figure 3.7: Realisation of the quantum switch in our framework through a quantum circuit.

and  $\hat{V}$  is as given in (3.5). From here, we will assume that all parties have two possible measurements, and each measurement has two possible outcomes. Hence,  $x, y, z$  and  $a, b, c$  will be bits,  $x, a$  denoting Alice's input and output,  $y, b$  denoting Bob's input and output and  $c, z$  denoting Charlie's input and output. For Alice and Bob, an input of 0 will instruct them to measure in the computational basis and an input of 1 will instruct them to measure in the Fourier basis (The basis composed of the states  $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ ). Charlie will have the same instructions in spirit, however his computational basis states will be  $|1\rangle$  for the outcome 0 and  $|2\rangle$  for the outcome 1 (in the case that Charlie's input is 0) and his Fourier basis will be composed of states  $|\pm\rangle = \frac{1}{\sqrt{2}}(|1\rangle \pm |2\rangle)$  (in the case his input is 1). This technicality is due to the control  $|0\rangle_C$  state being reserved for a do nothing instruction, which is something that Charlie is not interested in measuring. The unitary describing such a procedure at the  $\hat{V}$  stage can be given by unitaries of the form;

$$(3.61) \quad V_{S_1, R}(x_1 = 1) = |+\rangle\langle +|_{S_1} \otimes \mathbb{1}_{R_1} + |-\rangle\langle -|_{S_1} \otimes (|0\rangle\langle 1|_R + |1\rangle\langle 0|_R).$$

Consider the case where the input variables are  $x = 0, y = 1, z = 1$ . After some calculation, we find the state at the end of the protocol to be

$$(3.62) \quad \begin{aligned} &|3\rangle_c \left( |+\rangle_{S_1} \left( \frac{1}{2\sqrt{2}} |+\rangle_{S_2} + \frac{1}{4} |0\rangle_{S_2} \right) |000\rangle_R + |-\rangle_{S_1} \left( \frac{1}{4} |0\rangle_{S_2} - \frac{1}{2\sqrt{2}} |+\rangle_{S_2} \right) |001\rangle_R \right. \\ &+ |+\rangle_{S_1} \left( \frac{1}{2\sqrt{2}} |-\rangle_{S_2} + \frac{1}{4} |0\rangle_{S_2} \right) |010\rangle_R + |-\rangle_{S_1} \left( \frac{1}{4} |0\rangle_{S_2} - \frac{1}{2\sqrt{2}} |-\rangle_{S_2} \right) |011\rangle_R \\ &\left. + \frac{1}{4} |+\rangle_{S_1} |1\rangle_{S_2} |100\rangle_R + \frac{1}{4} |-\rangle_{S_1} |1\rangle_{S_2} |101\rangle_R - \frac{1}{4} |+\rangle_{S_1} |1\rangle_{S_2} |110\rangle_R - \frac{1}{4} |-\rangle_{S_1} |1\rangle_{S_2} |111\rangle_R \right) |111\rangle_F, \end{aligned}$$

where we have adopted the convention that  $|000\rangle_R = |a = 0, b = 0, c = 0\rangle_R$ . From this state, probabilities to obtain outcomes can be calculated using the Born rule. Let us focus on one of the terms to illustrate what we need to show, and leave the others for an extended exercise. The term associated to  $p(000|011)$  is a good candidate, since it notably contains interference from the two causal orders  $A \rightarrow B \rightarrow C$  and  $B \rightarrow A \rightarrow C$  (in contrast to say  $p(111|011)$ , which does not). Calculating,  $p(000|011) = \left| \frac{1}{2\sqrt{2}} |+\rangle_{S_1} |+\rangle_{S_2} + \frac{1}{4} |+\rangle_{S_1} |0\rangle_{S_2} \right|^2 = 5/16$ .

Imagine the probability associated to a ‘naive causal’ strategy, a closest classical sibling to the quantum switch  $p^{\text{nc}}(000|011)$ . From the set-up (3.59) of an equal superposition of  $|1\rangle_C$  and  $|2\rangle_C$ , one can argue that one half of the time, we are in the causal order  $A \rightarrow B \rightarrow C$ , Alice measures 0 with certainty and Bob, and Charlie have each a 50 : 50 chance to measure either 0 or 1. The other half of the time, we are in the causal order  $B \rightarrow A \rightarrow C$  all parties have a 50 : 50 chance to measure either 0 or 1 (since Bob’s measurement in the Fourier basis, which occurs first, now makes Alice completely uncertain of her outcome). Putting this all together we find  $p^{\text{nc}}(000|011) = \frac{1}{2} \times 1 \times \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 3/16 \neq p(000|011)$ .

Our results however, claim that we must be able to find at least one causal strategy which replicates the quantum  $p(000|011)$ . By direct substitution into equations (3.6) and (3.7), we find that:

$$\begin{aligned}
 p_1(l_1 = \text{Alice}|H_0) &= 1/2 \\
 p_1(l_1 = \text{Bob}|H_0) &= 1/2 \\
 p_2(l_2 = \text{Bob}|H_1 = (1, 0, 0)) &= 1 \\
 p_2(l_2 = \text{Alice}|H_1 = (2, 0, 1)) &= 1 \\
 p_3(l_3 = \text{Charlie}|H_2 = ((1, 0, 0), (2, 0, 1))) &= 1 \\
 p_3(l_3 = \text{Charlie}|H_2 = ((2, 0, 1), (1, 0, 0))) &= 1 \\
 p_1(a = 0|H_0, x = 0) &= 1 \\
 p_1(b = 0|H_0, y = 1) &= 1/2 \\
 p_2(a = 0|H_1 = (1, 0, 0), x = 0) &= 1/2 \\
 p_2(b = 0|H_1 = (2, 0, 1), y = 1) &= 1/2
 \end{aligned}$$

(3.63)

which are the same as the naive causal case. However, at the final stage we notice some differences. In particular we find

$$(3.64) \quad p_3(c = 0|H_2 = ((1, 0, 0), (2, 0, 1), z = 1)) = \frac{5}{6} \quad p_3(c = 0|H_2 = ((2, 0, 1), (1, 0, 0), z = 1)) = \frac{5}{6}.$$

Despite the ordering of the history for the classical protocol being different in these two cases, the quantum calculation given by (3.6) and (3.7) is the same for both (as it only depends on the flags raised and results obtained before Charlie measures). This leads to interference between the two causal orders of  $A$  and  $B$  in  $|\phi_{(l_3=\text{Charlie}, 3, H_2)}\rangle$ .

Putting it all together, the ‘alternative causal’ procedure which emulates the quantum experiment (the switch) therefore gives  $p^{\text{ac}}(000|011) = \frac{1}{2} \times \frac{1}{2} \times \frac{5}{6} + \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{5}{6} = 5/16$  as desired. Although we have focused on only one probability and strategy here, the same method can be used to generate a full classical strategy which replicates the quantum experiment for all input and output choices.

### 3.6 Equivalence Between Combined And Individual Controlled Lab Gates

In this section we show that the framework for quantum experiments is equivalent to considering any quantum circuit built up of standard unitary gates and controlled gates for individual laboratories, in terms of the probability distributions they can generate. The key idea involves showing how to construct individual controlled lab gates from  $V$ , and conversely how to construct the operation  $V$  from individual controlled lab gates.

To map any circuit involving individual controlled lab gates into our framework, we first space out the gates in the circuit, so that there is only one gate per time-step (this will increase the depth, which is not a resource we particularly care about). If an individual lab gate acts on only part of the system, we extend it such that it acts on the entire system, taking the action to be trivial (i.e. tensored with the identity) on any part of the system which was not initially included. We then replace individual controlled lab gates by a circuit fragment involving one use of  $V$ , using the approach described below. Finally we merge all unitary gates between instances of  $V$  into the unitaries  $U_t$ . This will lead to a circuit in our framework yielding exactly the same results as the original circuit. To go in the other direction, we simply replace each instance of  $V$  with its construction in terms of individual controlled lab gates.

Figures 3.8 and 3.9 show how to construct an individual controlled lab gate using  $V$ . Figure 3.10 shows how to construct  $V$  from individual controlled lab gates.

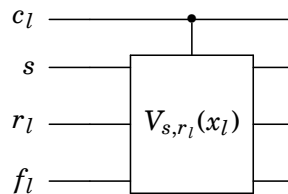


Figure 3.8: A controlled lab gate for an individual laboratory

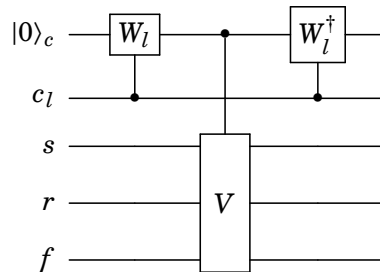


Figure 3.9: An equivalent circuit to the individual controlled lab gate above, built from a single instance of  $V$ , where  $W_l|0\rangle = |l\rangle$ . Note that the individual wires may represent composite subsystems rather than individual qubits.

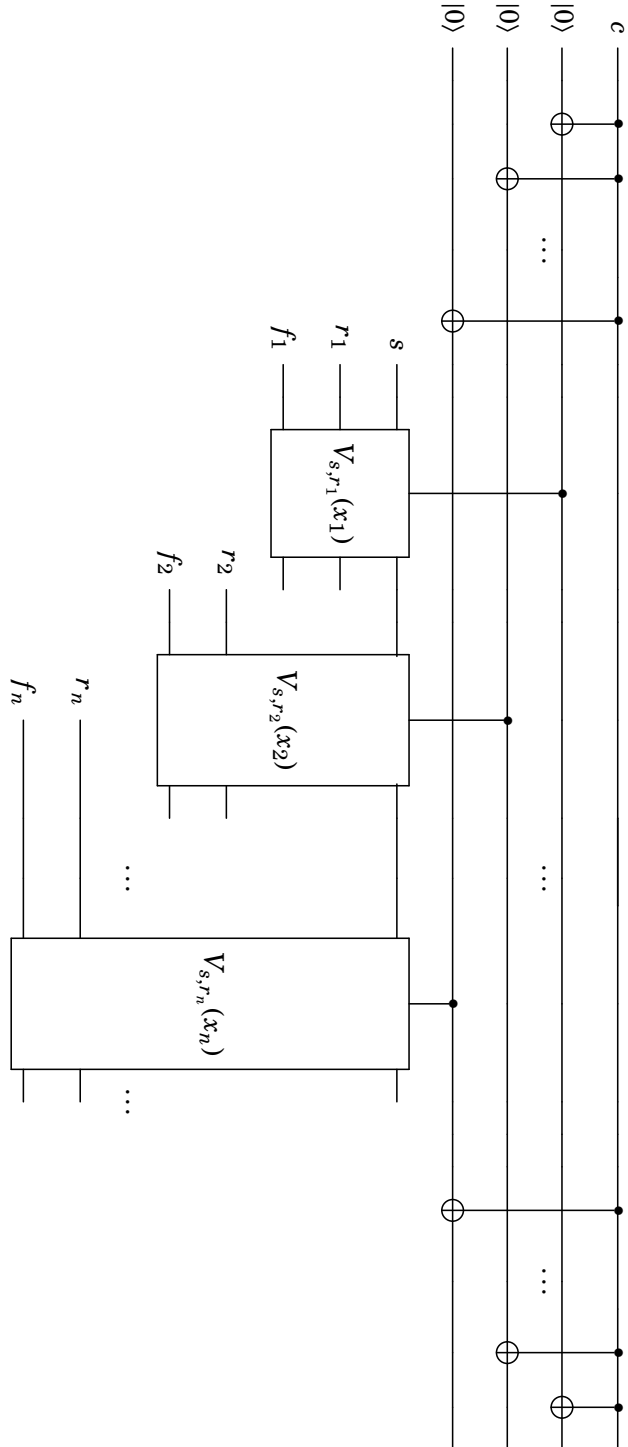


Figure 3.10:  $Y$ , built from individual controlled lab gates. The first and last  $CNOT$  gate are controlled from the state  $|1\rangle_c$ , the second and second from last are controlled from  $|2\rangle_c$ , and so on until the  $n$ 'th and  $n + 1$ 'th  $CNOT$ , which are controlled from state  $|n\rangle_c$ .

### 3.7 Discussions and Conclusions

In this chapter, we have examined what physically realisable superpositions of causal order of events mean in the light of the causal inequalities. Causal inequalities share many characteristic qualities with their more widely studied siblings, the Bell inequalities. They are both families of inequalities bounding convex sets of probability for one, and for another there is an operational manner in which to define them - as non-local games in the case of Bell inequalities and as causal games in the case of causal inequalities. It is through quantum theory that the analogy ends. In the case of non-locality, there is a distinction between the local polytope - the non-local correlations that can be achieved by sharing a classical system, and the quantum polytope - those non-local correlations achievable in quantum theory. It is a fact that we can prepare stronger correlations between space-like separated systems if we use quantum mechanics than if we use classical physics. The causal case, not so. We have shown in this chapter that causal inequalities are not violated in quantum theory, which echos the sentiment of non-local physics as just as an echo is - distorted. For quantum physics, despite containing causal non-separability in the interference between causal orders of events in space time, the polytope of achievable correlations matches the classical case exactly.

Through quantum control of causal order, and treating measurements as unitary processes, quantum theory permits superpositions of causal order and interference between them. Although our framework is very general, there so arise some natural criticisms that are interesting avenues for further research. Firstly, one key requirement is that each party only interacts with the state once. This is instantiated through the flag system. This is a normal requirement in set-up of causal inequalities and allows us to represent the experiment via  $p(\vec{a}|\vec{x})$ , with one output for every input. However, it would be interesting to investigate if lifting this restriction allows one to violate a causal inequality. It seems that it may, since given the power to erase results (in essence, to forget) adaptively based upon current information may provide the experimenter the power to trick himself. For an example of this, consider the example in the introduction based upon showing a recording of a post-selected experiment to a third party. The third party could be convinced, if they had no knowledge of post-selection, and they had faith in the randomising process, that information was being sent from future to the past. This is known to grant the power to violate causal inequalities. Similarly, it is known (and quite obvious) that allowing two-way communication can cause the violation of causal inequalities (I say obvious since any probability distribution can be created locally, or with enough two-way communication). However, fundamental (without cheating the causal model (3.1)) violations of causal inequalities in systems where quantum theory governs are impossible. We should note that if we want to take advantage of an adaptive protocol as described, then we require that the measurement results are stored in some part of the system, perhaps in one of the ancillas that is passed with the system during the protocol.

We have also, for natural reasons, assumed that the protocol takes a finite amount of time

steps. This seems natural given the framework. None-the-less, generalisation to both continuous time, and to an infinite amount of time-steps are both interesting avenues for further research. Generalisations to quantum field theory are also begging to be investigated, however the waters are muddied in this arena. Firstly, how does one rigorously define a closed, local laboratory for a quantum field? Next, how does one account for the quantum control over order, as in the circuit model? What is the fair causal inequality to assign to a given scenario? What about the vacuum state? We have in this work made an effort to include the zero, do nothing command so that a vacuum state is always present, when nothing else is. Although this is only a small step toward a field theoretic generalisation, the question is so interesting that we hope this work will form a basis for the next round of study.

Our framework also assumes standard quantum theory. If a theory changes significantly (perhaps to incorporate quantum gravity, or even just time dilation [181]) we might expect new possibilities for causal inequality violation. We hope that insights from this work will provide insight into these areas by making it clear what quantum theory can and cannot achieve.

The violation of a causal inequality is a theory independent fact, and fundamental violation of a causal inequality in an experiment would be one of the most exciting scientific discoveries of our century. It could hint at a physical scenario which was not well described by quantum theory. Not just that it was ill posed, or too complicated to calculate, - but that *could not* be described by quantum theory at a fundamental level. There has not been one of these scenario's since its conception - excluding of course the mystery of quantum gravity.

This thesis came with the promise of being driven by physical principles. In the quantum mechanical case, the principle of locality subtly fails us. However, the principle of causality, where the causal order of events is classical, holds.

## NON-CLASSICALLY CAUSAL CORRELATIONS WITHOUT BACKWARDS-IN-TIME SIGNALLING

This chapter is based off of the work ‘Non-classically causal correlations without backwards-in-time signalling’ [64], which appeared in the journal *Physical Review A*. It was completed at the University of Bristol in conjunction with my supervisor, Dr. Anthony J Short. In it, we revisit the no-backwards-in-time signaling (NBTS) scenario from the work ‘Exploring the limits of no-backwards-in-time signalling’ [198] and show that unlike the case of two parties, allowing three or more parties in the game does not force a classically causal distribution to arise. In short, local causality as an axiom does not necessarily enforce a causal description at the level of the global distribution to be permitted.

### 4.1 Introduction

What happens exactly to forbid messages arriving to us from the future? Surely, we cannot signal backwards in time in any sensible theory - paradoxes galore!? Quantum theory is a probabilistic theory, and it might surprise the reader to learn that this grants a remarkable amount of autonomy to the arrow of time. We shall see in this chapter that there exist theories that permit the future to effect the past, without permitting local backwards in time signalling.

As mentioned in the previous chapter, there has been a growing interest in the quantum information community surrounding indefinite causal order [204, 188, 178, 65, 180, 60]. The implementations of indefinite causal order are numerous, quantum switches [179, 178], post-selection [53, 194, 195, 196, 106], or (perhaps) a quantum theory of gravity [205, 206, 181] are all able to manifest this property.

One of the most general objects describing quantum theory in the absence of a pre-defined



causal order is the process matrix [188]. A process matrix is an object which colloquially can be understood to correspond to laboratories who obey quantum theory, but communicate without an explicit embedding in a time. We then imagine that the state can be thrown between laboratories of the experimenters without any regard to the arrow of time that we are used to! The state perhaps could travel between outputs of labs at a later ‘time’ to inputs of labs at an earlier ‘time’ and this can be leveraged to signal backwards to the past. This formalism has been shown to grant one the ability to violate the causal inequalities of the previous chapter [180, 199, 190, 207], and even so strong than that, it can be done in a classical (not quantum) manner by using a diagonal process matrix [64]. Using a process matrix to describe the connections between players in the GYNI or LGYNI game allows one to win the game perfectly! Operational quantum tasks also glean benefits by having access to a process matrix [208], so causal structure is a useful resource for information processing and computation. Digressing on that point, it is hard to imagine a computation that does not take advantage of a causal structure in some way. Suppose you have a ‘slider puzzle’ where the goal is to slide blocks (which are restricted in their motion by other blocks on the playing field) to reveal a message hidden in the centre of the board. Someone could give you ‘the solution’, but if they gave it to you in a jumbled causal order, you might still have a lot of work to do should you want to read the message. One is left either attempting to unscramble the jumbled up solution, or to just go ahead and solve the slider puzzle. There is no guarantee that the latter will be harder than the former. Sudoku is somewhat the same, although there it is less clear how a solution in a jumbled causal order effects you. Certainly though, if you talk through a jumbled causal order solution to a Sudoku master and were asked to explain the reasoning you applied when solving the problem, you may very well feel a hot sensation come under your collar for fear of being exposed as the cheat that you are!

Attempts to ascribe ‘physically realisable’ onto the process matrices is an ongoing area of active research, however key developments have been made [200, 202, 81]. One notes that although the process matrices do bear semblance to quantum theory, they are not derived from it. We shall see in this chapter how the two are connected. Our method for connecting the two is drawn as a generalisation of [209], and will ultimately depend upon the intuitions of Aharonov [53] and the two-time states.

The remainder of the chapter is structured as follows. In section 4.2 we explore a new way to do experiments, relying on the two-time states. Then, in section 4.3 we will explore a natural set of conditions obeyed by certain special types of experiments that naturally encode the intuition surrounding ‘no messages to the past’ but still without referencing clocks. In section 4.4 we will look at a subset of the two-time states, the linear two-time states and show that these are the ones which agree with the intuitive ideas of section 4.3 for all possible experiments. Then, in section 4.5 we find an explicit way to respect the NBTS conditions, but still violate a causal inequality. We will finish by discussing how one can generate the data using standard quantum theory. Finally we discuss our results in section 4.6.

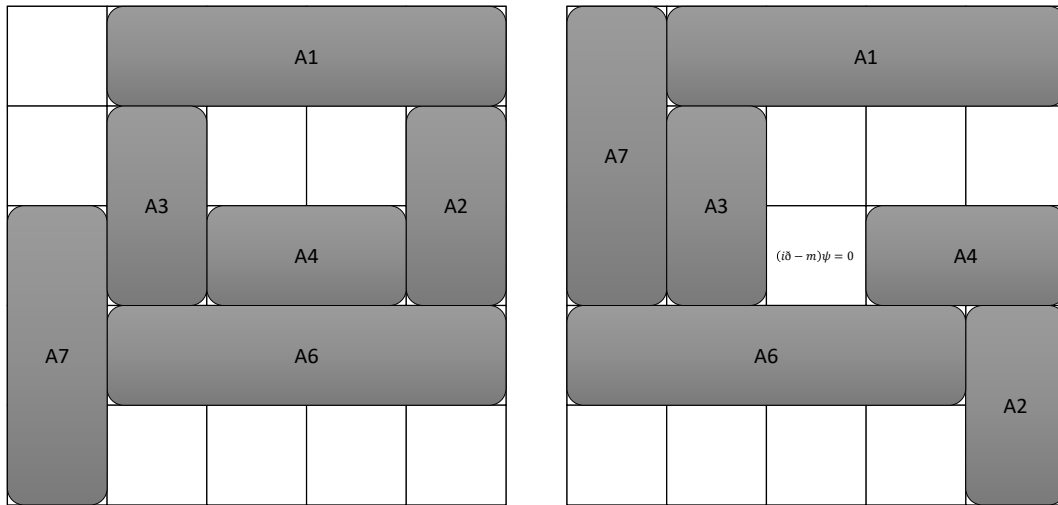


Figure 4.1: A slider puzzle, where the rules are that the blocks can only be moved in direction of their long axis, and the walls of the 5x5 grid are hard boundaries. Here is the solution: “A7, 2U; A6, 1L; A2, 2D; A4, 1R”. And here is a jumbled causal order solution: “A2, 2D; A4, 1R; A6, 1L; A7, 2U”. The jumbled causal order solution does not solve the problem (and actually, for this slider problem, there is only one move possible at each time-step, so there is no advantage at all in the jumbled order solution). The order in which you follow the steps in the solution is clearly a resource in the problem.

## 4.2 Two-Time States

We will expand first on the main tools with which we work, the ABL rule (1.26) as outlined in the introduction providing us with most of the insights needed. Most quantum experiments and algorithms involve a preparation, where a  $\rho$  is the initial condition for the set-up. The system is permitted to evolve (to probe ‘what is going on?’ - an experiment or ‘by design’ - an algorithm). After the evolution, a measurement is performed and hopefully, unless the measurement is maximally uninformative then we learn something about the state (in the experiment, or do some useful processing in the case of an algorithm). Since we know something, at least partially about the preparation, evolution and measurement, we can make statements that make inferences surrounding nature or some problem of interest. For instance in quantum theory, if we knew the preparation perfectly and we make enough measurement results, and we know that the evolution is closed to external environments, we could make the inference that quantum theory might be a unitary theory.

However, there is another way to do experiments, introduced in [53] by Aharonov and others. Since we typically pick an initial state in an experiment, instead of making a final measurement

we can fix a final state also (we could also consider the case where the initial state is an unknown but an initial measurement is made. This is trivial for the case of strong projective measurements (the state will collapse to an eigenvector of the projection) but for POVM measurements it is less clear). More specifically, at time  $t_1$  a state  $|\phi\rangle$  is prepared. In the time interval up to  $t_2$ , arbitrary experiments are performed with this state. At  $t_2$ , and observable  $\mathcal{A}$  is measured, which has a non degenerate eigenvectors  $|\psi\rangle$ . The experiment is considered a success if the outcome associated to  $|\psi\rangle$  is observed and a fail else. In the case of a failure, the statistics of the intermediate time experiment are discarded. We note that the sub-ensemble of statistics is in general quite different to the full ensemble of non-discarded statistics [196, 209]. Being a quantum procedure, in full generality, there is in general no way to know *ex-ante* that the post-selection will succeed. There are specific cases where one can, and these will have important implications later down the line.

An important distinction in the literature is between experimental post-selections described above, and fundamental post-selections. Fundamental post-selection is described by exactly the same mathematics as the experimental post-selection procedure outlined above, however it has quite different and profound implications. Fundamental post-selection occurs when nature is providing the final boundary condition and has been suggested to be a route to understanding the physics of black holes [210], or the fate of the universe [211].

We will now expand on the ABL formula in definition (1.26), and in doing so we will develop a language for talking about such experiments which we will refer to as the two-time language. We begin with a restatement of definition (1.26) for completeness.

**Definition 4.1.** *Suppose a state of a system  $A$  is prepared at time  $t_1$  in  $|\phi\rangle$ . Suppose that the Hamiltonian on the system is trivial for simplicity. The system is subjected to a detailed measurement, described by a POVM with Kraus operators  $\hat{E}_a = \sum_{k,l} c_{k,l}^a |k\rangle\langle l|$ , where detailed is taken to mean that there is only one Kraus operator associated to each possible measurement outcome. Thus, the normalisation of the Kraus operators is  $\sum_a \hat{E}_a^\dagger \hat{E}_a = \mathbb{1}$ . Then the system at a later time  $t_2$  is post-selected into the state  $|\psi\rangle$ . The probability to observe outcome  $a$  is given by the ABL rule:*

$$(4.1) \quad p(a) = \frac{|\langle\psi|\hat{E}_a|\phi\rangle|^2}{\sum_{a'} |\langle\psi|\hat{E}_{a'}|\phi\rangle|^2}.$$

From here on in, in this chapter, we will adopt the two-time language. Formally, we take the state space of the preparer and post-selector to be  $\mathcal{H}_{\mathcal{A}_o} \otimes \mathcal{H}^{\mathcal{A}_i}$ , where  $\mathcal{H}_{\mathcal{A}_o}$  is the space where the post-selection occurs, and we adopt the convention that states in this space are denoted by bra-vectors  $\langle\psi|_{\mathcal{A}_o}$ . The space  $\mathcal{H}^{\mathcal{A}_i}$  is the state space where the prepared state lives, and as usual with states denoted as ket-vectors  $|\phi\rangle^{\mathcal{A}_i}$ . The reasons for this notation will become clear. The structure is made yet more explicit by associating to a two-time language version of a state and a

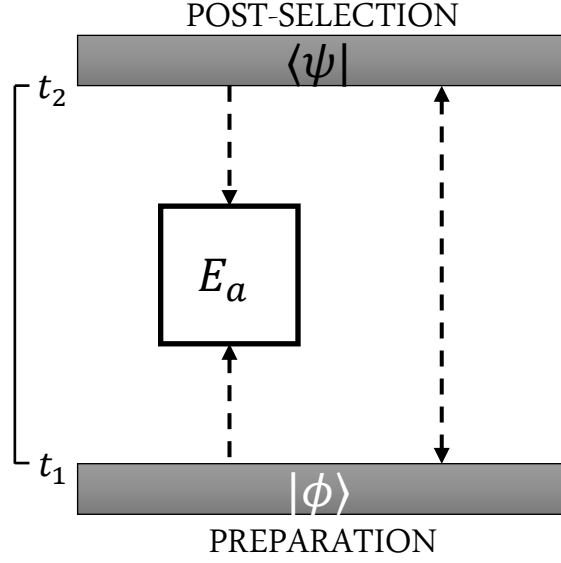


Figure 4.2: A diagram of a quantum process involving post-selection. At some early time  $t = t_1$ , the system is prepared in a state  $|\phi\rangle$ . An experiment is conducted and, at a later time  $t = t_2$ , the system is post-selected into the state  $|\psi\rangle$ .

Kraus operator in the following way,

$$(4.2) \quad \begin{aligned} \Phi &= \langle \psi |_{\mathcal{A}_o} \otimes |\phi\rangle_{\mathcal{A}_i} \\ E_a &= \sum_{k,l} c_{k,l} |k\rangle_{\mathcal{A}_o} \otimes \langle l |_{\mathcal{A}_i} \end{aligned}$$

This notation does a couple of things for us. Firstly, there is now a clear symmetry between states and measurements - they are obtained from one another by the dual, or dagger operation. Furthermore, input and output spaces are now clearly separated. This will allow us the freedom to connect Hilbert spaces in exotic manners, through the dot operation;  $\bullet$ . The dot operation is defined to take vectors and dual vectors in the same Hilbert space and combine them through the usual scalar product, and take vectors and dual vectors from different Hilbert spaces and combine them as a tensor product. For example, we can re-write the amplitude from equation (4.1) as  $\Phi \bullet E_a$ .

**Definition 4.2.** *More specifically, we can define the dot as the linear operator satisfying  $|i\rangle_{\mathcal{A}} \bullet \langle j|_{\mathcal{A}} = \langle j|_{\mathcal{A}} \bullet |i\rangle_{\mathcal{A}} = \langle i|j\rangle$  and  $|i\rangle_{\mathcal{A}} \bullet \langle j|_{\mathcal{B}} = \langle j|_{\mathcal{B}} \bullet |i\rangle_{\mathcal{A}} = |i\rangle_{\mathcal{A}} \otimes \langle j|_{\mathcal{B}}$ .*

Using the dot operation and the two-time language, we can re-write the ABL rule (4.1) as

$$(4.3) \quad p(a) = \frac{|\Phi \bullet E_a|^2}{\sum_{a'} |\Phi \bullet E_{a'}|^2}.$$

A general pure state in the two-time language can be given by

$$(4.4) \quad \Psi = \sum_{i,j} \alpha_{i,j} \langle i |_{\mathcal{A}_o} \otimes |j\rangle_{\mathcal{A}_i}$$

Which can be viewed as fundamentally post-selected quantum state, or a state which is experimentally pre and post-selected, possibly making use of entangled ancillas. By generalising further, one can consider mixtures of two-time states in the following manner. Preparing an ensemble of pure states  $\{\Phi_r = \sum_{i,j} \alpha_{i,j}^r |i\rangle_{\mathcal{A}_o} \otimes |j\rangle_{\mathcal{A}_i}\}$  with probabilities  $p_r$ , then it is possible to write down the ABL rule for the ensemble as

$$(4.5) \quad p(a) = \frac{\sum_r p_r |\Phi_r \bullet E_a|^2}{\sum_{a'} \sum_{r'} p_{r'} |\Phi \bullet E_{a'}|^2}.$$

In standard quantum theory, the density matrix is the object which accounts for combinations of superposition and classical randomness simultaneously at a single time. The standard way to obtain a density matrix from a (pure) quantum state is  $\rho = |\psi\rangle \otimes \langle\psi|$ . For a two-time generalisation, we write the density vector as  $\Psi \otimes \Psi^\dagger$ , such that for the pure state in (4.4)

$$(4.6) \quad \begin{aligned} \Psi \otimes \Psi^\dagger &\in \mathcal{H}_{\mathcal{A}_o} \otimes \mathcal{H}^{\mathcal{A}_i} \otimes \mathcal{H}^{\mathcal{A}_o^\dagger} \otimes \mathcal{H}_{\mathcal{A}_i^\dagger} \\ \Psi \otimes \Psi^\dagger &= \sum_{i,j,n,m} \alpha_{i,j} \alpha_{n,m}^* |i\rangle_{\mathcal{A}_o} \otimes |j\rangle_{\mathcal{A}_i} \otimes \langle m|_{\mathcal{A}_o^\dagger} \otimes |n\rangle_{\mathcal{A}_i^\dagger} \end{aligned}$$

with a general density vector given as a convex combination over the ensemble of pure density vectors,  $\eta = \sum_r p_r \Psi_r \otimes \Psi_r^\dagger$ . Analogously, we can construct a ‘Kraus density vector’ for measurement procedures. Dropping the requirement that a single Kraus operator corresponds to each measurement outcome and instead allowing for coarse grained measurements, the density vector corresponding to the outcome  $a$  is a sum over all the Kraus operators pertaining to that outcome. If  $\hat{E}_a^\mu$  denotes the Kraus operator in standard quantum theory,  $a$  the outcome and  $\mu$  an index running over all operators corresponding to  $a$ , then the Kraus density vector for the measurement is given by

$$(4.7) \quad J_a = \sum_\mu E_a^\mu \otimes E_a^{\mu\dagger} \in \mathcal{H}_{\mathcal{A}_i} \otimes \mathcal{H}^{\mathcal{A}_i^\dagger} \otimes \mathcal{H}^{\mathcal{A}_o} \otimes \mathcal{H}_{\mathcal{A}_o^\dagger}.$$

We note that  $J_a$  contains all the information about the dynamics of the measurement (and so the post measurement state conditioned on outcomes) and the outcome probabilities. This distinction is important since it makes the object  $J_a$  distinct from the POVM element  $\sum_\mu \hat{E}_a^\mu \hat{E}_a^{\mu\dagger}$ . In the standard formalism, normalisation is a condition on measurement operators (dating back ultimately to dutch book arguments for the normalisation of probability [212]) achieved by demanding that  $\sum_a \sum_\mu \hat{E}_a^{\mu\dagger} \hat{E}_a^\mu = \mathbb{1}$ . In the two-time language, this condition is concisely written as

$$(4.8) \quad \mathbb{1}_{\mathcal{A}_o^\dagger} \bullet \sum_a J_a = \mathbb{1}_{\mathcal{A}_i^\dagger}.$$

Where we have introduced the identity vector for the input,  $\mathbb{1}_{\mathcal{A}_i^\dagger} = \sum_i |i\rangle_{\mathcal{A}_i^\dagger} \otimes \langle i|_{\mathcal{A}_i}$  and correspondingly for the output system,  $\mathbb{1}_{\mathcal{A}_o^\dagger} = \sum_i |i\rangle_{\mathcal{A}_o^\dagger} \otimes \langle i|_{\mathcal{A}_o}$ . We refer to this property as future identity preserving.

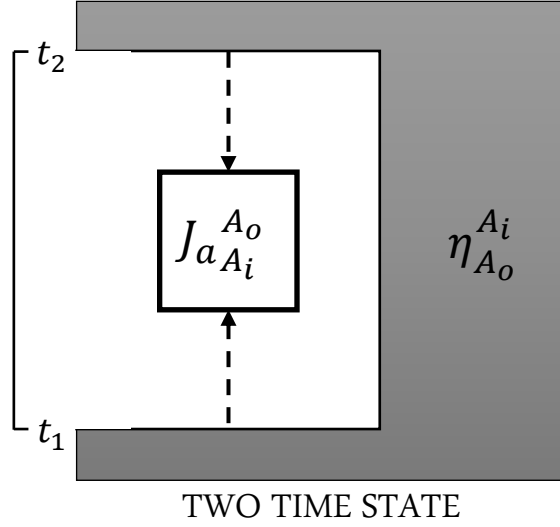


Figure 4.3: A two-time state experiment.

**Definition 4.3.** *Future identity preserving maps carry the property that*

$$(4.9) \quad \uparrow_{\mathcal{A}_o}^{\mathcal{A}_i^\dagger} \bullet J_{A_i}^{A_o} = \uparrow_{\mathcal{A}_i}^{\mathcal{A}_o^\dagger}$$

where  $\bullet J_{A_i}^{A_o} = \sum_a \bullet J_{a A_i}^{A_o}$

By further simplifying notation, such that the ‘full Hilbert space’  $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}^{\mathcal{A}^\dagger}$  is written in capital Roman format as  $\mathcal{H}_A$ , the equation for outcome statistics in the mixed case (4.5) can be written using the  $\bullet$  operation as

$$(4.10) \quad p(a) = \frac{\eta_{A_o}^{A_i} \bullet J_{a A_i}^{A_o}}{\sum_{a'} \eta_{A_o}^{A_i} \bullet J_{a' A_i}^{A_o}}$$

Contractions between indices, through the dot operation, adopt a similar convention to Einstein summation notation where repeated indices on the upstairs and the downstairs are contracted. This is well and good for experiments generating a single outcome  $a$ , however we know that far more general scenarios than this are possible. What about the case of two - or more parties?

In fact the two-time language is rather general and captures these scenarios well. For example, the case of a bipartite probability distribution between A (Alice) and B (Bob) can be written down in a similar manner,

$$(4.11) \quad p(a, b) = \frac{\eta_{A_o B_o}^{A_i B_i} \bullet (J_{a A_i}^{A_o} \otimes K_{b B_i}^{B_o})}{\sum_{a'} \eta_{A_o B_o}^{A_i B_i} \bullet (J_{a' A_i}^{A_o} \otimes K_{b B_i}^{B_o})}$$

Which is the situation depicted in figure 4.4b. We can also consider a conditional distribution

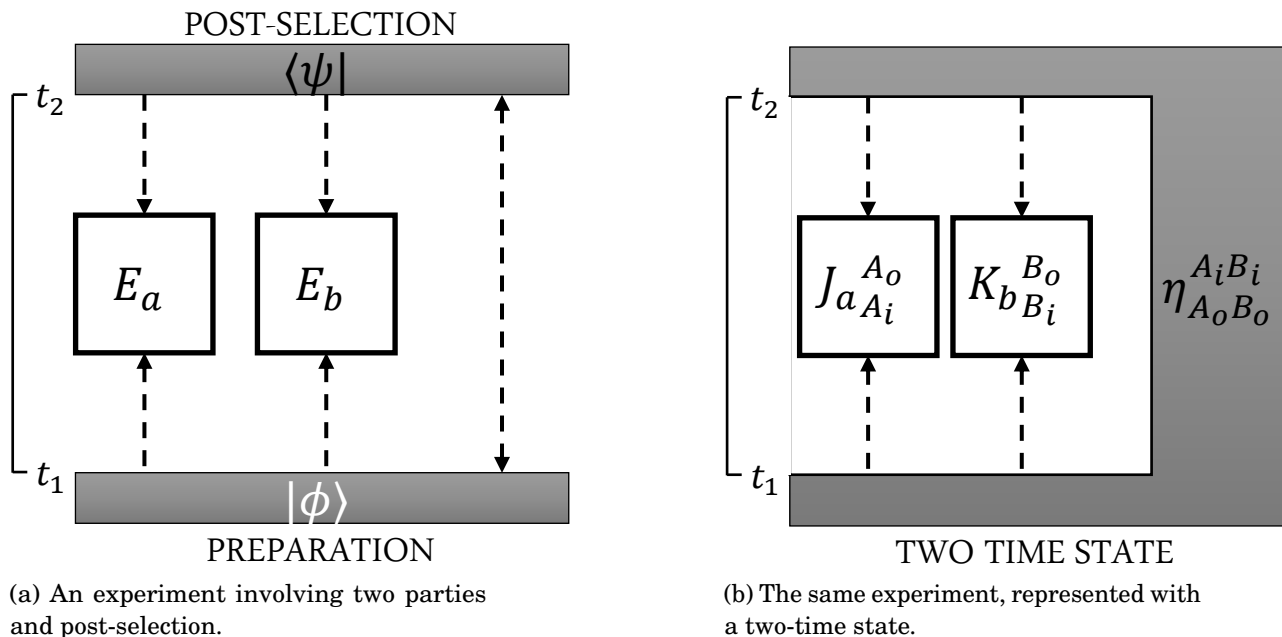


Figure 4.4: Two experiments, equivalently described by standard quantum theory and post-selection or a two-time state.

which has input dependence on some random variables  $x, y$  as,

$$(4.12) \quad p(a, b|x, y) = \frac{\eta_{A_o B_o}^{A_i B_i} \bullet (J_{a A_i}^{x A_o} \otimes K_{b B_i}^{y B_o})}{\sum_{a', b'} \eta_{A_o B_o}^{A_i B_i} \bullet (J_{a' A_i}^{x A_o} \otimes K_{b' B_i}^{y B_o})},$$

with  $J_{a A_i}^{x A_o}$  and  $K_{b B_i}^{y B_o}$  the measurement operators associated to outcomes  $a$  and  $b$ , upon the instruction to measure in a basis specified by  $x$  and  $y$ . Furthermore, causal order can be captured in this framework. Suppose we are in the no-input scenario, and it is certain that Alice will perform her measurement of the state before Bob performs his. In general, there will be some interaction between the system and an environment in the intermediary time, but in the interest of keeping things simple we ignore this in this work and assuming that the system is just passed from pre-selection to Alice, directly to Bob and then directly to the post-selection. In this case the statistics of such an experiment are captured by

$$(4.13) \quad p(a, b) = \frac{\eta_{B_o}^{A_i} \bullet (J_{a A_i}^{A_o} \otimes K_{b B_i}^{B_o}) \bullet \mathbb{1}_{A_o}^{B_i}}{\sum_{a', b'} \eta_{B_o}^{A_i} \bullet (J_{a' A_i}^{A_o} \otimes K_{b' B_i}^{B_o}) \bullet \mathbb{1}_{A_o}^{B_i}},$$

and is depicted in figure 4.5. We can also add even more parties in the intuitive fashion. Important for this work, is the case of three parties. For a conditional tripartite two-time state, the statistics are calculated by

$$(4.14) \quad p(a, b, c|x, y, z) = \frac{\eta_{A_o B_o C_o}^{A_i B_i C_i} \bullet (J_{a A_i}^{x A_o} \otimes K_{b B_i}^{y B_o} \otimes L_{c C_i}^{z C_o})}{\sum_{a', b', c'} \eta_{A_o B_o C_o}^{A_i B_i C_i} \bullet (J_{a' A_i}^{x A_o} \otimes K_{b' B_i}^{y B_o} \otimes L_{c' C_i}^{z C_o})},$$

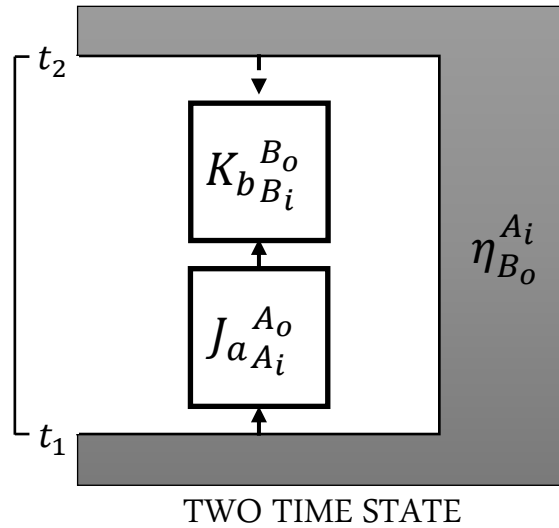


Figure 4.5: Depiction of the experiment described by equation (4.13).

### 4.3 The NBTS Paradigm

We now outline a natural paradigm for causal investigation, from which we will be able to outline a local principle for a direction of time and investigate the possible constraints (and possibilities) on global distributions that manifest from these local properties. The no-backwards-in-time signalling regime is a regime for local laboratories to conduct experiments, obeying some strict but highly intuitive methodology:

1. A system enters a laboratory.
2. A fixed measurement of the received system is made, and a record made of the classical data.
3. A classical input is received. This can be thought of as being gifted by an external agent, produced from a random number generator, or representing the free will on behalf of the agent.
4. A transformation of the system is made, which may depend upon the received input.
5. The system exits the laboratory

This prescription is the NBTS paradigm. If the causal order is flowing normally we expect that the outputs of the experiment done in step 2 cannot depend upon the inputs in step 3.

**Definition 4.4.** *Considering a single party, the marginal will read*

$$(4.15) \quad p(a|x) = p(a).$$

*This condition is the NBTS condition.*



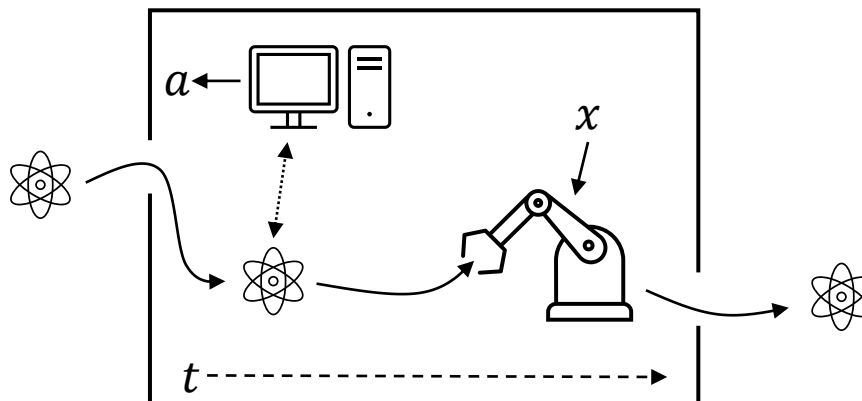


Figure 4.6: Pictorial representation of a laboratory in the NBTS paradigm. At some point, a system enters into a laboratory. A measurement of that system is made, and then the system gets transformed (based on some random variable  $x$ ). The system then leaves the lab.

At no point in this prescription do we mention time, although it is perfectly reasonable to give the experimenter inside the labs a clock if we so wish. The problem with this is that synchronisation of clocks is not an *in antecessum* given, so in the multi partyed case experiments between laboratories would have to be conducted to synchronise the clocks. For two parties repeating the experiment in succession this is no problem, since one is sure that in the next passing of state between labs where the state will end up, and indeed where it has just come from (the NBTS condition holding also implies that the experimenter will be sure the state is not sent backward to his own input, and there is only one choice left - that the state must be sent to the party that is not you!). For three or more parties, obscured is the option for the causal order to be so well defined - it could be adaptive based on the measurement outcomes. Charlie is no longer sure if he in synchronising with Alice and then *responsible* for Bob's synchronicity in every round. We will return to this point.

In a recent article [198] experiments within the NBTS were (theoretically) performed in a quantum mechanical setting in which each party performs a fixed measurement followed by a chosen transformation of the state. In this scenario, a variety of options are considered for the relative timings of two laboratories in the bipartite case. Most generally, it was shown that for two parties, a quantum experiment involving post-selection satisfied the NBTS conditions if and only if it can be represented by a process matrix [198, 209] (as we shall see, equivalently, a linear two-time state).

The correlations obtained in the two-party NBTS scenario were limited to the classically causal set (see previous chapter). This holds for two parties, despite the existence of consistent non-classical probabilities and causal indefiniteness and causal non-separability. It is therefore an interesting question whether process matrices always lead to classical correlations in the NBTS scenario.

We now will take a moment to outline the NBTS conditions for three parties because they

will be important for what follows, before resuming our study of the relationship between the ABL rule, process matrices and linear two-time states.

Consider three parties (Alice, Bob and Charlie), now with each party obeying the NBTS prescription. Alice gets input  $x$  and outputs  $a$ , Bob and Charlie get inputs  $y$  and  $z$  and output  $b$  and  $c$  respectively. The overall statistics of the experiment are well described by the global distribution  $p(a, b, c|x, y, z)$ .

**Definition 4.5.** *The NBTS paradigm ensures that marginalising over Bob and Charlie, to obtain the distribution for Alice, should be independent of her input*

$$(4.16) \quad p_A(a|x, y, z) = \sum_{b, c} p(a, b, c|x, y, z) = p_A(a|y, z).$$

*By demanding that the same hold for Bob and Charlie we find*

$$(4.17) \quad p_B(b|x, y, z) = \sum_{a, c} p(a, b, c|x, y, z) = p_B(b|x, z)$$

$$(4.18) \quad p_C(c|x, y, z) = \sum_{a, b} p(a, b, c|x, y, z) = p_C(c|x, y)$$

*for all  $x, y, z, a, b, c$ .*

We are able to describe the permitted probability distributions as a convex polytope in probability space. The variables  $p(a, b, c|x, y, z)$  act as our co-ordinates and the value of  $p$  a scalar on that space. Certain points in this space are forbidden probabilities - either ruled out because they do not satisfy normalisation, positivity or, in the case of NBTS distributions, not obeying the NBTS conditions. For three parties, assuming the  $x, y, z, a, b, c$  are binary, the situation is described by 64 co-ordinates specified by the  $2^6$  possible combinations of  $x, y, z, a, b, c$ . Respecting positivity and normalisation means that the following are satisfied

$$(4.19) \quad 0 \leq p(a, b, c|x, y, z) \leq 1 \quad \sum_{a, b, c} p(a, b, c|x, y, z) = 1.$$

We also require that the probability distribution obeys the NBTS conditions (4.16) (4.17). These conditions specify a convex set which we refer to as the NBTS-polytope.

A classically causal polytope sits within the NBTS polytope which contains the distributions which can be achieved using a classical strategy of the kind outlined in the last chapter. The deterministic strategies sit at the vertices of such a convex polytope and hence a polytope can be generated in vertex representation by considering all the deterministic strategies.

Let us enumerate some of these strategies for the tripartite scenario. Suppose Alice goes first. This means her output can only ever be a constant bit, it occurs before any input. There is a freedom now to choose the causal order, based upon the input. For example, it may be the case that then  $x = 0$  the system is next given to Bob and when  $x = 1$  the system is given next to

Charlie. Such a strategy is described by the formula

$$(4.20) \quad p(a, b, c|0, y, z) = \begin{cases} 1 & \text{if } a = \alpha, b = \beta, c = \delta y \oplus \gamma \\ 0 & \text{otherwise} \end{cases}$$

$$(4.21) \quad p(a, b, c|1, y, z) = \begin{cases} 1 & \text{if } a = \alpha, c = \lambda, b = \nu z \oplus \mu \\ 0 & \text{otherwise.} \end{cases}$$

where  $\alpha, \beta, \gamma, \delta, \lambda, \mu, \nu$  are bits and  $\oplus$  denotes addition modulo 2. We are able to then construct the full polytope by enumerating all strategies of this form (and the ones where Charlie and Bob are both first to act) and constructing the convex hull. We find by direct computation that there are 560 such vertices, once redundancies are removed. A facet representation is computationally hard to determine, and for the scope of the current discussion irrelevant thus has been omitted.

Finally we also define a quantum polytope, which is analogous to the polytope of the same name in the studies of non-signalling. Strategies within the quantum polytope are those that are allowed to use process matrices outside the local laboratories and standard quantum theory inside the laboratories. In [209] it was demonstrated that the quantum and classical polytopes for the two party scenario are identical, and hence there are no operational differences between a quantum scenario and one demanding no-backwards-in-time signalling. Could this be another manifestation of the work of the previous chapter? In other words, despite there being plenty of exotic ways to wire together laboratories in the quantum scenario, do the local NBTS conditions demand that the global distribution is essentially classical? Is the NBTS scenario enough to ensure such a causal description?

We will show this is, surprisingly, not the case. Before that however, we should spend some time discussing a subset of the two-time states, the linear two-time states.

## 4.4 Linear Two-Time States

A subset of the possible two-time states naturally encode the idea of no-backwards-in-time-signalling required in the previous section. We call this subset the linear two-time states. Hence, if we can describe a generalised post-selected experiment with a linear two-time state, then we know the experiment will obey no-backwards-in-time-signalling.

An important first step in uncovering what set this corresponds to is found in theorem 1 of [198], which we restate here for completeness.

**Theorem 4.1.** *States for a single party satisfying no-backwards-in-time-signalling correspond to states without post-selection, or states with trivial post-selection. If the two-time state  $\eta = \eta_{A_0}^{A_i}$  obeys the NBTS conditions for all outcome statistics then*

$$(4.22) \quad \eta_{A_0}^{A_i} = \rho^{A_i} \otimes \mathbb{1}_{A_0}$$

where  $\rho$  is a valid density matrix.

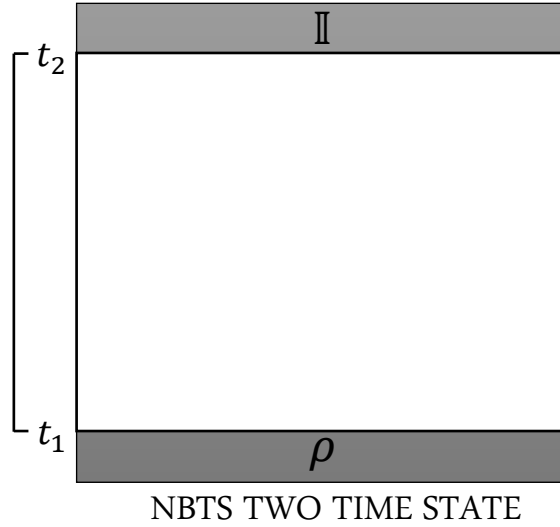


Figure 4.7: A state for a single party that ensures the NBTS conditions are met, irrespective of the measurements that are made in the middle. This corresponds to a preparation of a quantum state, and a trivial post-selection.

A vector  $\eta_{A_o B_o C_o}^{A_i B_i C_i}$  is a linear two-time state if the ABL rule (4.14) reduces to

$$(4.23) \quad p(a, b, c | x, y, z) = \eta_{A_o B_o C_o}^{A_i B_i C_i} \bullet (J_{a, A_i}^x \otimes K_{b, B_i}^y \otimes L_{c, C_i}^z),$$

or equivalently if the summation

$$(4.24) \quad \sum_{a', b', c'} \eta_{A_o B_o C_o}^{A_i B_i C_i} \bullet (J_{a', A_i}^x \otimes K_{b', B_i}^y \otimes L_{c', C_i}^z)$$

is equal to one. We note here that (4.23) is a linear function of both the state and the measurements, hence the name.

**Theorem 4.2.** *The linear two-time states correspond exactly to the process matrices, for any number of parties.*

In order to prove theorem (4.2), we will begin at the definition of probabilities in the process matrix formalism given in [188]. Note here there are no conditionals in order to simplify the notation, these can be added onto the measurement operators without altering the proof structure.

We begin with the probability rule for process matrices.

$$\begin{aligned}
P(a_1, \dots, a_N) &= \\
&= \text{Tr} \left[ W^{A_1^1 A_2^1 \dots A_1^N A_2^N} \left( \bigotimes_{i=1}^N M_{a_i}^{A_i^1 A_i^2} \right) \right] \\
&= \text{Tr} \left[ \sum_{\mu_1 \dots \mu_N} W^{A_1^1 A_2^1 \dots A_1^N A_2^N} \left( \bigotimes_{i=1}^N \left[ (\mathbb{1} \otimes \hat{E}_{\mu_i, a_i}^{[i]}) |\Phi^+\rangle \langle \Phi^+| (\mathbb{1} \otimes \hat{E}_{\mu_i, a_i}^{[i]\dagger}) \right]^T \right) \right] \\
&= \text{Tr} \left[ \sum_{\mu_1 \dots \mu_N} (W^{A_1^1 A_2^1 \dots A_1^N A_2^N})^T \left( \bigotimes_{i=1}^N \left[ (\mathbb{1} \otimes \hat{E}_{\mu_i, a_i}^{[i]}) |\Phi^+\rangle \langle \Phi^+| (\mathbb{1} \otimes \hat{E}_{\mu_i, a_i}^{[i]\dagger}) \right] \right) \right] \\
&= \text{Tr} \left[ \sum_{\substack{k_1 \dots k_{2N} \\ l_1 \dots l_{2N}}} \sum_{\mu_1 \dots \mu_N} w_{k_1 \dots k_{2N}, l_1 \dots l_{2N}} |l_1 \dots l_{2N}\rangle \langle k_1 \dots k_{2N}| \left( \bigotimes_{i=1}^N \left[ (\mathbb{1} \otimes \hat{E}_{\mu_i, a_i}^{[i]}) |\Phi^+\rangle \langle \Phi^+| (\mathbb{1} \otimes \hat{E}_{\mu_i, a_i}^{[i]\dagger}) \right] \right) \right] \\
&= \sum_{\substack{k_1 \dots k_{2N} \\ l_1 \dots l_{2N}}} \sum_{\mu_1 \dots \mu_N} w_{k_1 \dots k_{2N}, l_1 \dots l_{2N}} \left( \prod_{i=1}^N \text{Tr} \left[ |l_{2i-1} l_{2i}\rangle \langle k_{2i-1} k_{2i}| (\mathbb{1} \otimes \hat{E}_{\mu_i, a_i}^{[i]}) \sum_{t_i, u_i} |t_i t_i\rangle \langle u_i u_i| (\mathbb{1} \otimes \hat{E}_{\mu_i, a_i}^{[i]\dagger}) \right] \right) \\
&= \sum_{k_1 \dots k_{2N}} \sum_{l_1 \dots l_{2N}} \sum_{\mu_1 \dots \mu_N} w_{k_1 \dots k_{2N}, l_1 \dots l_{2N}} \left( \prod_{i=1}^N \langle k_{2i} | \hat{E}_{\mu_i, a_i}^{[i]} |k_{2i-1}\rangle \langle l_{2i-1} | \hat{E}_{\mu_i, a_i}^{[i]\dagger} |l_{2i}\rangle \right) \\
&= \left( \sum_{k_1 \dots k_{2N}} \sum_{l_1 \dots l_{2N}} w_{k_1 \dots k_{2N}, l_1 \dots l_{2N}} \mathcal{A}_2^1 \langle k_2 | \otimes |k_1\rangle^{\mathcal{A}_1^1} \otimes \mathcal{A}_1^{1\dagger} \langle l_1 | \otimes |l_2\rangle^{\mathcal{A}_2^{1\dagger}} \otimes \dots \right. \\
&\quad \left. \dots \otimes \mathcal{A}_2^N \langle k_{2N} | \otimes |k_{2N-1}\rangle^{\mathcal{A}_1^N} \otimes \mathcal{A}_1^{N\dagger} \langle l_{2N-1} | \otimes |l_{2N}\rangle^{\mathcal{A}_2^{N\dagger}} \right) \cdot \left( \sum_{\mu_1} \hat{E}_{\mu_1, a_1}^{[1]} \otimes \hat{E}_{\mu_1, a_1}^{[1]\dagger} \otimes \dots \otimes \sum_{\mu_N} \hat{E}_{\mu_N, a_N}^{[N]} \otimes \hat{E}_{\mu_N, a_N}^{[N]\dagger} \right) \\
&= \eta_W \bullet (J_{a_1}^{[1]} \otimes \dots \otimes J_{a_N}^{[N]})
\end{aligned}
\tag{4.25}$$

which is equivalent to the probability rule for linear two-time states (4.23), generalised to multiple parties each performing a channel  $J_{a_i}^{[i]}$ . Furthermore, note that  $\eta$  is ‘positive’ (in the sense required to be a valid two-time state [196]) if and only if  $W$  is a positive operator. Following [209, 64] we observe a natural isomorphism between process matrices and the linear two-time states. The isomorphism between  $W$  and  $\eta$  is obtained by expressing the former as a vector in the basis chosen for the transpose and flipping bras and kets on the output spaces - one can expect this since the process formalism makes heavy use of the Choi isomorphism [213] and so to move back to a picture of state, it will be necessary to take a transpose.

Theorem (4.1) can be shown in the multiple party case to generalise to the following theorem

**Theorem 4.3.** *A general two-time state for any number of parties satisfies the NBTS conditions for all possible measurements if and only if it is linear.*

Which we now prove. To see the ‘only if’ part of the theorem, that a linear two-time state satisfies the NBTS conditions, note that it is by the results of theorem (4.2) it is equivalent to a process matrix. The process matrices are defined so that individual laboratories satisfy standard

quantum theory without post-selection, and so cannot lead to backwards in time signalling. Alternatively, we construct a multiparty generalisation of equation (30) in [209] and use the argument in theorem 2 of [198].

We have therefore still to prove that satisfying the NBTS conditions implies that a general pre and post-selected state can be represented by a linear two-time state. First, consider the marginal state over the first party, given by

$$(4.26) \quad \eta_{A_1} = (\mathcal{J}^{[2]} \otimes \dots \otimes \mathcal{J}^{[N]}) \bullet \eta$$

where  $\mathcal{J}^{[k]}$  is the channel for the  $k$ 'th party summed over their output, which may depend on their input  $x_k$ , and otherwise is totally arbitrary. Theorem (4.1) then says that this state must be proportional to a state which has identity over the output corresponding to no post-selection. Thus

$$(4.27) \quad \begin{aligned} (\mathcal{J}^{[1]} \otimes \mathcal{J}^{[2]} \otimes \dots \otimes \mathcal{J}^{[N]}) \bullet \eta &= (\mathcal{J}^{[1]})_{A_1^1}^{A_2^1} \bullet (\rho^{A_1^1} \otimes \mathbb{1}_{A_2^1}) \\ &= \mathbb{1}_{A_1^1} \bullet \rho^{A_1^1} \end{aligned}$$

where we have used the future identity preserving property of trace-preserving channels,  $\mathbb{1}_B \bullet \mathcal{J}_A^B = \mathbb{1}_A$ . Note that (4.27) is independent of  $\mathcal{J}^{[1]}$  and so does not depend on  $x_1$ . Repeating these arguments for all the other parties,  $(\mathcal{J}^{[1]} \otimes \dots \otimes \mathcal{J}^{[N]}) \bullet \eta$  must be a constant independent of the channels  $\mathcal{J}^{[1]}$  through to  $\mathcal{J}^{[N]}$  (and hence independent of all of the inputs  $x_1, \dots, x_N$ ). By appropriately normalising  $\eta$  (which doesn't effect the physics of two-time states [55]) we can obtain

$$(4.28) \quad (\mathcal{J}^{[1]} \otimes \dots \otimes \mathcal{J}^{[N]}) \bullet \eta = 1$$

for all choices of channels  $\mathcal{J}^{[1]}$  to  $\mathcal{J}^{[N]}$ . From the general rule for probabilities in the two-time state formalism,

$$(4.29) \quad \begin{aligned} p(a_1, \dots, a_N) &= \frac{(\mathcal{J}_{a_1}^{[1]} \otimes \dots \otimes \mathcal{J}_{a_N}^{[N]}) \bullet \eta}{(\mathcal{J}^{[1]} \otimes \dots \otimes \mathcal{J}^{[N]}) \bullet \eta} \\ &= (\mathcal{J}_{a_1}^{[1]} \otimes \dots \otimes \mathcal{J}_{a_N}^{[N]}) \bullet \eta \end{aligned}$$

which is the probability rule given in (4.23), and thus  $\eta$  is a linear two-time state if and only if it obeys the NBTS conditions.

## 4.5 Procedure

It is not clear how the quantum and classically causal polytopes can be different. Indeed, in the case of bipartite linear two-time states they are not. Here we show how they can be different, by giving an explicit example of a linear two-time state which allows the violation of causal

inequalities. Since causal inequalities are not violated by the classically causal set by definition, we are done. To reiterate, here we will construct an explicit protocol which both respects the NBTS conditions, but none the less violates a causal inequality and as such does not sit within the classically causal set. We will use the same wiring of laboratories as first discovered by Bäümeler *et al.* in [65], but a new protocol that respects the NBTS paradigm.

Alice, Bob and Charlie are sat within closed laboratories. They are assumed to have no knowledge about the outside world, or how they are connected to one another. They each perform the following protocol: at some point a qubit enters their lab, and they first measure it in the computational basis, and record the outcome of their measurement. Then, they receive a classical input bit. If this input is 0, they pass the quantum system out of the door of their labs. If they instead get 1, they first apply the  $\hat{X} = |0\rangle\langle 1| + |1\rangle\langle 0|$  operation and then pass it out. The action of Alice, whose input space is  $A_i$  and output is  $A_o$  and who received input  $x$  and obtained output  $a$  in her measurement can be represented by

$$(4.30) \quad M(a|x)_{A_i}^{A_o} = |a \oplus x\rangle_{\mathcal{A}_o} \otimes_{\mathcal{A}_i} \langle a| \otimes |a\rangle_{\mathcal{A}_i^\dagger} \otimes_{\mathcal{A}_o^\dagger} \langle a \oplus x|,$$

with Bob and Charlie represented by

$$(4.31) \quad M(b|y)_{B_i}^{B_o} = |a \oplus x\rangle_{\mathcal{B}_o} \otimes_{\mathcal{B}_i} \langle a| \otimes |a\rangle_{\mathcal{B}_i^\dagger} \otimes_{\mathcal{B}_o^\dagger} \langle a \oplus x|,$$

$$(4.32) \quad M(c|z)_{C_i}^{C_o} = |a \oplus x\rangle_{\mathcal{C}_o} \otimes_{\mathcal{C}_i} \langle a| \otimes |a\rangle_{\mathcal{C}_i^\dagger} \otimes_{\mathcal{C}_o^\dagger} \langle a \oplus x|.$$

Outside the laboratories, and unbeknown to Alice, Bob and Charlie the laboratories are connected in a rather odd fashion. As we do not require that standard quantum theory apply here, we can allow strange connections between the labs. The labs are connected as depicted in figure 4.8, where the output of Alice is connected to the input of Bob, the output of Bob connected to the input of Charlie and the output of Charlie connected to the input of Alice! Clearly this is impossible to implement in standard quantum theory, but we avoid any paradoxes - at least locally, through obeying the NBTS conditions. The NBTS conditions are satisfied since with probability 1/2 the labs are connected through  $\hat{Z}$  basis measurements, and with probability 1/2 the labs are connected through  $\hat{Z}$  basis measurements followed by bit-flip gates. Colloquially, since there are an odd number of parties and Alice has no idea which of these two options is the ‘real’ one, the quantum system that she passes out of her lab becomes randomised as it arrives back at her input. She cannot signal to her own input, which are exactly what the NBTS conditions enforce. Any action she takes to manipulate the state will be invisible in what she is able to calculate at this point in the experiment, which is an approximation (due to finite rounds) to  $p_A(a|x, y, z)$ .

This analogy corresponds exactly to the case where Bob and Charlie perform identity gates, however it is a nice interpretation of the mathematics that moreover works for all channel combinations, which we will show momentarily in a more formal argument.

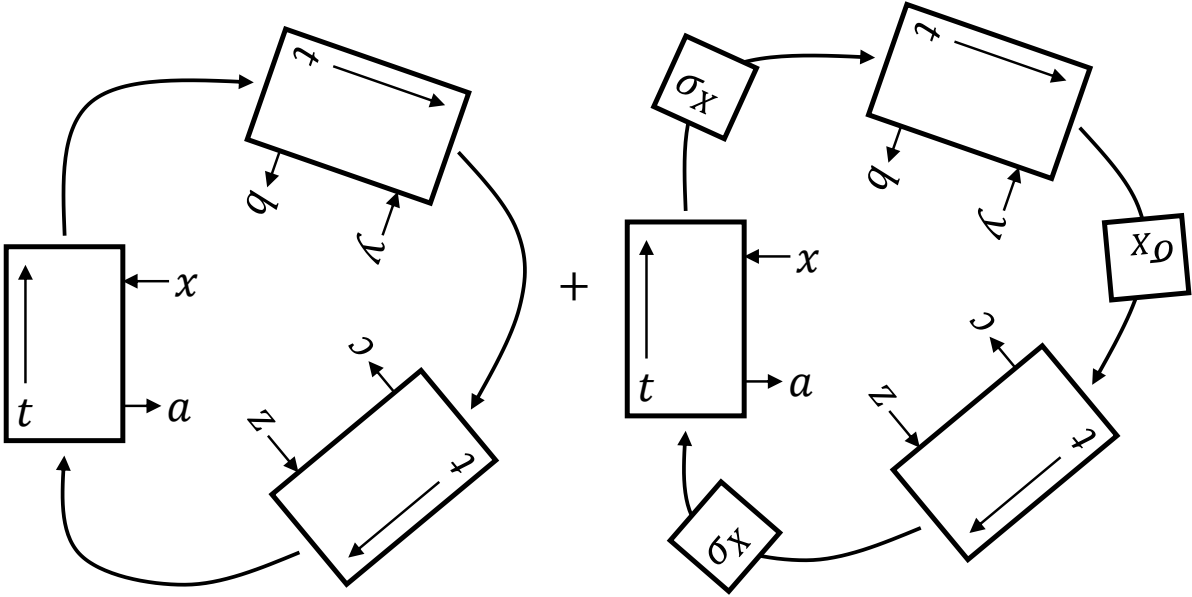


Figure 4.8: A picture of the protocol that achieves the NBTS, but causal inequality violating statistics described in the main text. The black lines connecting laboratories represent channels with a  $z$ -basis measurement. The gates labelled  $X$  are flip gates. Notice that in the local frame of each laboratory the output is produced before the input is received.

More formally, the situation outside the laboratories is described by a linear two-time state. This state is given by

$$(4.33) \quad \eta = \frac{1}{2}(M_{A_o}^{B_i} \otimes M_{B_o}^{C_i} \otimes M_{C_o}^{A_i} + \bar{M}_{A_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i} \otimes \bar{M}_{C_o}^{A_i})$$

where the operators  $M$  and  $\bar{M}$  are given as

$$(4.34) \quad M_A^B = M(0|0)_A^B + M(1|0)_A^B$$

$$(4.35) \quad \bar{M}_A^B = M(0|1)_A^B + M(1|1)_A^B.$$

In order to continue with this discussion we should first prove that this set-up does indeed respect the NBTS conditions. This amounts, as we have seen, to verifying that  $\eta$  is linear two-time state, and so to showing that the relation

$$(4.36) \quad (J \otimes K \otimes L) \bullet \eta = 1$$

holds for all trace-preserving channels  $J$ ,  $K$ , and  $L$  that Alice, Bob and Charlie could perform.

Observe that  $\eta$  contains a  $z$ -basis measurement in every link between parties. This means it is sufficient to consider only classical stochastic channels for the parties. Why so? Because we observe the fact that any quantum channel that is sandwiched between two  $z$ -measurements is stochastic, and superposition is destroyed. Consider a quantum channel with Kraus operators



$A_k$ , ‘sandwiched’ between two  $z$  basis measurements;

$$(4.37) \quad M_z[\mathcal{E}[M_z(\rho)]] = \sum_{i,j,k} |j\rangle\langle j|A_k|i\rangle\langle i|\rho|i\rangle\langle i|A_k^\dagger|j\rangle\langle j| = \sum_{i,j} q_{j|i} |j\rangle\langle i|\rho|i\rangle\langle j|$$

where  $q_{j|i}$  is equal to

$$(4.38) \quad q_{j|i} = \sum_k \langle j|A_k|i\rangle\langle i|A_k^\dagger|j\rangle.$$

$q_{j|i}$  is clearly positive and real, being the magnitude of a complex number. We can identify (4.37) with a classical stochastic map if the numbers  $q_{j|i}$  can be identified with probabilities. Lets do that quickly,

$$(4.39) \quad \sum_i q_{j|i} = \sum_{k,i} \langle j|A_k|i\rangle\langle i|A_k^\dagger|j\rangle = \sum_k \langle j|A_k A_k^\dagger|j\rangle = \langle j|\mathbb{1}|j\rangle = 1.$$

Thus, any quantum channel for qubits that gets sandwiched between two  $z$ -basis measurements can be simulated by a stochastic channel.

In the case of classical bits, there are a set of four channels that are extreme points of the convex set of classical channels - the identity, the flip or  $X$  gate, and the throw away and replace with 0 and 1 channels. All other channels are convex mixtures of these four channels. If (4.36) holds for these it will hold for any convex combination of classical channels by linearity. Moreover, since all of the channels  $J, K, L$  are future identity preserving, instead of checking all  $4^3$  channel combinations, we can focus in on one party of interest and check only the 4 extremal channels for this party. Without loss of generality, consider Alice as the party of interest. As a first case, let  $J$  be the identity channel. Then consider

$$(4.40) \quad (\mathbb{1}_{A_o}^{A_o} \otimes K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet \frac{1}{2} (M_{A_o}^{B_i} \otimes M_{B_o}^{C_i} \otimes M_{C_o}^{A_i} + \bar{M}_{A_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i} \otimes \bar{M}_{C_o}^{A_i}).$$

By using the identity on Alice to connect and contract the indices through the dot operator this becomes

$$(4.41) \quad (K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet \frac{1}{2} (M_{A_o}^{B_i} \bullet M_{C_o}^{A_o} \otimes M_{B_o}^{C_i} + \bar{M}_{A_o}^{B_i} \bullet \bar{M}_{C_o}^{A_o} \otimes \bar{M}_{B_o}^{C_i}).$$

We have the relations that  $M_A^B \bullet M_B^C = M_A^C$  and  $\bar{M}_A^B \bullet \bar{M}_B^C = \bar{M}_A^C$  which gives

$$(4.42) \quad (K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet \frac{1}{2} (M_{C_o}^{B_i} \otimes M_{B_o}^{C_i} + M_{C_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i})$$

$$(4.43) \quad = (K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet (M_{C_o}^{B_i} \otimes \frac{1}{2} (M_{B_o}^{C_i} + \bar{M}_{B_o}^{C_i})).$$

We note that  $\frac{1}{2} (M_{B_o}^{C_i} + \bar{M}_{B_o}^{C_i})$  is the operation throw away and replace with the maximally mixed state, which has the operational form  $\mathbb{1}_{B_o} \otimes \frac{1}{2} \mathbb{1}^{C_i}$ . All the remaining channels are by definition trace preserving - or future identity preserving in the two-time language - they satisfy  $\mathbb{1}_B \bullet J_A^B = \mathbb{1}_A$ , so

we can move the identity through the state giving

$$\begin{aligned}
& \frac{1}{2}(K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet (M_{C_o}^{B_i} \otimes \mathbb{1}_{B_o} \otimes \mathbb{1}^{C_i}) \\
&= \frac{1}{2}(\mathbb{1}_{B_i} \otimes L_{C_i}^{C_o}) \bullet (M_{C_o}^{B_i} \otimes \mathbb{1}^{C_i}) \\
&= \frac{1}{2}L_{C_i}^{C_o} \bullet (\mathbb{1}_{C_o} \otimes \mathbb{1}^{C_i}) \\
&= \frac{1}{2}\mathbb{1}_{C_i} \bullet \mathbb{1}^{C_i} \\
(4.44) \quad &= 1.
\end{aligned}$$

We should also confirm this when Alice applies the flip channel;

$$(4.45) \quad X_A^B = |0\rangle^{\mathcal{B}} \otimes_{\mathcal{A}} \langle 1| \otimes |1\rangle^{\mathcal{A}^\dagger} \otimes_{\mathcal{B}^\dagger} \langle 0| + |1\rangle^{\mathcal{B}} \otimes_{\mathcal{A}} \langle 0| \otimes |0\rangle^{\mathcal{A}^\dagger} \otimes_{\mathcal{B}^\dagger} \langle 1|$$

which gives

$$(4.46) \quad (X_{A_i}^{A_o} \otimes K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet \frac{1}{2}(M_{A_o}^{B_i} \otimes M_{B_o}^{C_i} \otimes M_{C_o}^{A_i} + \bar{M}_{A_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i} \otimes \bar{M}_{C_o}^{A_i}).$$

Now we have the relations that  $X_A^B \bullet M_B^C = \bar{M}_A^C$  and  $X_A^B \bullet \bar{M}_B^C = M_A^C$ . Then we get

$$\begin{aligned}
& (K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet \frac{1}{2}(M_{B_o}^{C_i} \otimes \bar{M}_{A_i}^{B_i} \bullet M_{C_o}^{A_i} + \bar{M}_{B_o}^{C_i} \otimes M_{A_i}^{B_i} \bullet \bar{M}_{C_o}^{A_i}) \\
&= (K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet \frac{1}{2}(M_{B_o}^{C_i} \otimes \bar{M}_{B_o}^{B_i} + \bar{M}_{B_o}^{C_i} \otimes \bar{M}_{C_o}^{B_i}) \\
&= (K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet \left(\frac{1}{2}(M_{B_o}^{C_i} + \bar{M}_{B_o}^{C_i}) \otimes \bar{M}_{C_o}^{B_i}\right) \\
&= \frac{1}{2}(K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet (\mathbb{1}_{B_o} \otimes \mathbb{1}^{C_i} \otimes \bar{M}_{C_o}^{B_i}) \\
(4.47) \quad &= 1.
\end{aligned}$$

By playing the same trick with the two other channel types we can recover similar results. For instance the throw away and replace with zero channel has the operational form

$$(4.48) \quad \mathbb{1}_{A_i} \otimes |0\rangle^{\mathcal{A}_o} \otimes_{\mathcal{A}_o^\dagger} \langle 0|.$$

The method here can be visualised easily - if we have a circular structure of future preserving channels, we can always move identity 'backwards' around the circle, until it 'hits' the prepared

state and gives 1. In particular

$$\begin{aligned}
& (\mathbb{1}_{A_i} \otimes |0\rangle_{\mathcal{A}_o} \otimes_{\mathcal{A}_o^\dagger} \langle 0| \otimes K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet \frac{1}{2} (M_{A_o}^{B_i} \otimes M_{B_o}^{C_i} \otimes M_{C_o}^{A_i} + \bar{M}_{A_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i} \otimes \bar{M}_{C_o}^{A_i}) \\
&= (|0\rangle_{\mathcal{A}_o} \otimes_{\mathcal{A}_o^\dagger} \langle 0| \otimes K_{B_i}^{B_o} \otimes L_{C_i}^{C_o}) \bullet \frac{1}{2} (M_{A_o}^{B_i} \otimes M_{B_o}^{C_i} \otimes \mathbb{1}_{C_o} + \bar{M}_{A_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i} \otimes \mathbb{1}_{C_o}) \\
&= (|0\rangle_{\mathcal{A}_o} \otimes_{\mathcal{A}_o^\dagger} \langle 0| \otimes K_{B_i}^{B_o} \otimes \mathbb{1}_{C_i}) \bullet \frac{1}{2} (M_{A_o}^{B_i} \otimes M_{B_o}^{C_i} + \bar{M}_{A_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i}) \\
&= (|0\rangle_{\mathcal{A}_o} \otimes_{\mathcal{A}_o^\dagger} \langle 0| \otimes K_{B_i}^{B_o}) \bullet \frac{1}{2} (M_{A_o}^{B_i} \otimes \mathbb{1}_{B_o} + \bar{M}_{A_o}^{B_i} \otimes \mathbb{1}_{B_o}) \\
&= (|0\rangle_{\mathcal{A}_o} \otimes_{\mathcal{A}_o^\dagger} \langle 0| \otimes \mathbb{1}_{B_i}) \bullet \frac{1}{2} (M_{A_o}^{B_i} + \bar{M}_{A_o}^{B_i}) \\
&= (|0\rangle_{\mathcal{A}_o} \otimes_{\mathcal{A}_o^\dagger} \langle 0|) \bullet \mathbb{1}_{A_o} \\
(4.49) \quad &= 1.
\end{aligned}$$

Clearly the above holds also for the throw away and replace with 1 channel. Hence, the state is a linear two-time state for all four extreme classical trace-preserving channels, and thus  $\eta$  is a linear two-time state.

Being a linear two-time state, our protocol satisfies the NBTS conditions, for all possible measurement and transformations Alice, Bob and Charlie can make. We now calculate the measurement probabilities for our protocol. We can use formula (4.23) to calculate the probabilities in this setting;

$$(4.50) \quad p(a, b, c|x, y, z) = (M(a|x)_{A_i}^{A_o} \otimes M(b|y)_{B_i}^{B_o} \otimes M(c|z)_{C_i}^{C_o}) \bullet \eta.$$

After some calculating, we can arrive at the compact relation for probabilities

$$(4.51) \quad p(a, b, c|x, y, z) = \frac{1}{2} (\delta_{b, a \oplus x} \delta_{c, b \oplus y} \delta_{a, c \oplus z} + \delta_{b, a \oplus \bar{x}} \delta_{c, b \oplus \bar{y}} \delta_{a, c \oplus \bar{z}})$$

where we use a bar above a bit to represent its logical negation (e.g.  $\bar{x} = x \oplus 1$ ). This generates the probabilities in table 4.1. The equation (4.51) itself carries some relations that we can use to

$p(a, b, c x, y, z)$		$a, b, c$							
		000	001	010	100	011	101	110	111
$x, y, z$	000	1/2	0	0	0	0	0	0	1/2
	001	0	0	1/2	0	0	1/2	0	0
	010	0	0	0	1/2	1/2	0	0	0
	100	0	1/2	0	0	0	0	1/2	0
	011	0	1/2	0	0	0	0	1/2	0
	101	0	0	0	1/2	1/2	0	0	0
	110	0	0	1/2	0	0	1/2	0	0
	111	1/2	0	0	0	0	0	0	1/2

Table 4.1: Table of probabilities generated by the protocol.

verify that no classically causal strategy could replicate this relation. Other alternatives could

make use of the formulation of a ‘causal witness’ [191], or a linear programming tool such as Irs [214] on the explicit polytope constructed in the previous section. Here, we favour analytic analysis since in this case it is simple, however in a noisy model one is expected to favour the witness method.

Namely, we have

$$(4.52) \quad \begin{aligned} p(a, b, c|x, y, z) &= p(\bar{a}, \bar{b}, \bar{c}|x, y, z) = p(a, b, c|\bar{x}, \bar{y}, \bar{z}) \\ &= p(b, c, a|y, z, x) = p(a, \bar{b}, c|x, y, \bar{z}). \end{aligned}$$

Let  $p_i$  be some deterministic strategy that could contribute to any affine combination of classical strategies that makes up  $p$ , i.e.  $p = \sum_i \lambda_i p_i$ , with  $0 < \lambda_i \leq 1$ . Then, take a particular choice of  $a, b, c, x, y, z$  such that

$$(4.53) \quad p_i(a, b, c|x, y, z) > 0,$$

Given that  $p(a, b, c|x, y, z) \geq \lambda_i p_i(a, b, c|x, y, z)$ , this is only possible if  $p(a, b, c|x, y, z) = 0.5$ , which we deduce by searching and comparing against the only non-zero value in the probability table. This also implies

$$(4.54) \quad p(a, \bar{b}, c|x, y, \bar{z}) = 0.5 \text{ and } p(\bar{a}, b, \bar{c}|x, y, \bar{z}) = 0.5.$$

Hence, using normalisation

$$(4.55) \quad p(a, b, c|x, y, \bar{z}) = 0.$$

The cyclic symmetry may be used to derive similar relations for the logical not of the other inputs, so that

$$(4.56) \quad p(a, \bar{b}, c|x, \bar{y}, z) = 0 \quad \text{and} \quad p(a, b, c|\bar{x}, y, z) = 0.$$

The deterministic classically causal strategy  $p_i$  has some party that goes last when the inputs are  $x, y, z$ . The probability distribution must be the same when the last party’s input is flipped (as it occurs after all outputs). Hence  $p_i(a, b, c|x, y, z) > 0$  implies

$$(4.57) \quad p_i(a, b, c|\bar{x}, y, z) + p_i(a, b, c|x, \bar{y}, z) + p_i(a, b, c|x, y, \bar{z}) > 0,$$

as one of the three terms must be positive. This further implies

$$(4.58) \quad p(a, b, c|\bar{x}, y, z) + p(a, b, c|x, \bar{y}, z) + p(a, b, c|x, y, \bar{z}) > 0,$$

which creates a contradiction with (4.55) and (4.56). Hence we cannot write  $p$  as a mixture of classically causal strategies.

Interestingly, although an effectively classical process which we could have conducted using a coin, the states being heads and tails. This point is actually not only an extreme point of the

classical set but also extreme in the NBTS set. The distribution in Table 4.1 can be seen to satisfy the positivity inequalities and normalisation equalities (Equations (4.19)). We can check using lrs [214] that the distribution in Table 4.1 saturates a total of 64 of the positivity and NBTS inequality constraints. In a 64 dimensional space, if we have saturated this number of inequalities it means that this probability distribution is the unique point satisfying all of these constraints. Hence, the only way for this point to be a mixture of points is if some of these points in probability space either violate the NBTS conditions or positivity (and normalisation). This method uses points outside of the NBTS polytope, and hence we confirm here that distribution (4.1) is an extreme point of the NBTS set.

It is perhaps not remarkable, in light of the last chapter, here that the extreme point can be achieved with a state and strategy that are both ‘classical’ - taken to mean that they are in some classical analogue of the process matrix set and not that causal loops are present in classical physics. Quantum theory gives no extra power over classical physics in terms of causal inequality violation, so the results of this section further strengthen the message that the interplay of quantum theory with causality is at least as subtle as the relationship of classical physics with causality.

### 4.5.1 Methods For Generating The Data

How can we simulate the above process? It will be a useful exercise and a demonstrative example of how exotic causality can be simulated with post-selection.

The first obvious thing to try is to use a closed-time-like curve. This will of course generate the correct statistics, as can be seen in figure 4.9b, by probabilistically picking (using classical randomness) whether or not to apply the  $\hat{X}$  gates during the time between the state being in the laboratories with  $p = 1/2$ , and running the protocol in the  $A \rightarrow B \rightarrow C$  causal order, and utilising the closed-time-like curve to pass the state from the output of the final operation at Charlie back to the input of Alice, we do indeed generate the correct statistics. There is a slight hitch, to the best of our knowledge closed-time-like curves are fantasy objects in the mathematical description of general relativity. Can we get these results by using post-selection?

The answer is yes we can. What we need to do is this. We first prepare a maximally entangled state. We take one half of this state, and apply the protocol to it, in the Alice  $\rightarrow$  Bob  $\rightarrow$  Charlie causal order, choosing whether or not to apply the  $\hat{X}$  gates between each lab with probability  $1/2$ . We then post select upon the maximally entangled state. Letting  $A_i, E$  be the Hilbert space associated to the input Bell pair,  $C, D$  be the Hilbert space associated to the post-selection upon a Bell pair, the final  $\hat{z}$  channel from the output of Charlie ( $C_o$ ) to the input of the post-selection.

The operators describing such a post selection, and  $\hat{z}$  channel are

$$\begin{aligned}\hat{p}s_{A_i E}^{CD} &= |\phi^+\rangle_{A_i E} \langle \phi^+ |^{A_i^\dagger E^\dagger} \otimes |\phi^+\rangle_{C^\dagger D^\dagger} \langle \phi^+ |^{CD} \otimes \mathbb{1}_D^E \otimes \mathbb{1}_{E^\dagger}^{D^\dagger} \\ \hat{z}_{C_o}^{C_o} &= |0\rangle_C \langle 0|^{C_o} \otimes |0\rangle_{C_o^\dagger} \langle 0|^{C_o^\dagger} + |1\rangle_C \langle 1|^{C_o} \otimes |1\rangle_{C_o^\dagger} \langle 1|^{C_o^\dagger}\end{aligned}$$

(4.59)

So that through the dot  $\bullet$  operation,

$$\begin{aligned}\hat{p}s_{A_i E}^{CD} \bullet \hat{z}_{C_o}^{C_o} &= |\phi^+\rangle_{A_i E} \langle \phi^+ |^{A_i^\dagger E^\dagger} \otimes |\phi^+\rangle_{C^\dagger D^\dagger} \langle \phi^+ |^{CD} \otimes \mathbb{1}_D^E \otimes \mathbb{1}_{E^\dagger}^{D^\dagger} \bullet (|0\rangle_C \langle 0|^{C_o} \otimes |0\rangle_{C_o^\dagger} \langle 0|^{C_o^\dagger} + |1\rangle_C \langle 1|^{C_o} \otimes |1\rangle_{C_o^\dagger} \langle 1|^{C_o^\dagger}) \\ &= |\phi^+\rangle_{A_i E} \langle \phi^+ |^{A_i^\dagger E^\dagger} \otimes (|0\rangle_{C_o^\dagger} \langle 0|_{D^\dagger} \langle 0|^{C_o} \langle 0|^D + |1\rangle_{C_o^\dagger} \langle 1|_{D^\dagger} \langle 1|^{C_o} \langle 1|^D) \\ &= |0\rangle_{A_i} \langle 0|^{C_o} \otimes |0\rangle_{C_o^\dagger} \langle 0|^{A_i^\dagger} + |1\rangle_a \langle 1|^{C_o} \otimes |1\rangle_{C_o^\dagger} \langle 1|^{A_i^\dagger}\end{aligned}$$

(4.60)

$$= \hat{z}_{A_i}^{C_o}$$

We teleport the output of the  $\hat{z}$  channel to the input at Alice. Hence, by pre and post-selecting on the Bell state  $|\phi^+\rangle$ , we create a  $\hat{z}$  channel from the output of Charlie to the input of Alice. A similar calculation shows that we can also create the  $\hat{z}x$  channel from the output of B to the input at A,

$$\begin{aligned}\hat{p}s_{A_i E}^{CD} \bullet \hat{z}x_{C_o}^{C_o} &= |\phi^+\rangle_{A_i E} \langle \phi^+ |^{A_i^\dagger E^\dagger} \otimes |\phi^+\rangle_{C^\dagger D^\dagger} \langle \phi^+ |^{CD} \otimes \mathbb{1}_D^E \mathbb{1}_{E^\dagger}^{D^\dagger} \bullet |0\rangle_C \langle 1|^{C_o} \otimes |1\rangle_{C_o^\dagger} \langle 0|^{C_o^\dagger} + |1\rangle_C \langle 0|^{C_o} \otimes |0\rangle_{C_o^\dagger} \langle 1|^{C_o^\dagger} \\ &= |0\rangle_{A_i} \langle 1|^B \otimes |1\rangle_{C_o^\dagger} \langle 0|^{A_i^\dagger} + |1\rangle_a \langle 0|^{C_o} \otimes |0\rangle_{C_o^\dagger} \langle 1|^{A_i^\dagger}\end{aligned}$$

(4.61)

$$= \hat{z}x_{A_i}^{C_o}$$

Letting the channel from A to B be well described by the operator

$$(4.62) \quad (M_{A_o}^{B_i} \otimes M_{B_o}^{C_i} + \bar{M}_{A_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i}) \bullet (M(a|x)_{A_i}^{A_o} \otimes M(b|y)_{B_i}^{B_o} \otimes M(c|z)_{C_i}^{C_o})$$

By mixing the process used to generate  $\hat{z}x_{A_i}^{C_o}$  and  $\hat{z}_{A_i}^{C_o}$  with one-half probability, we do indeed create the desired two-time state  $\eta$ ,

$$\begin{aligned}(4.63) \quad & (M_{A_o}^{B_i} \otimes M_{B_o}^{C_i} \otimes \frac{1}{2} \hat{p}s_{A_i E}^{CD} \bullet \hat{z}_{C_o}^{C_o} + \bar{M}_{A_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i} + \frac{1}{2} \hat{p}s_{A_i E}^{CD} \bullet \hat{z}x_{C_o}^{C_o}) \\ &= \frac{1}{2} (M_{A_o}^{B_i} \otimes M_{B_o}^{C_i} \otimes M_{C_o}^{A_i} + \bar{M}_{A_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i} \otimes \bar{M}_{C_o}^{A_i}) = \eta\end{aligned}$$

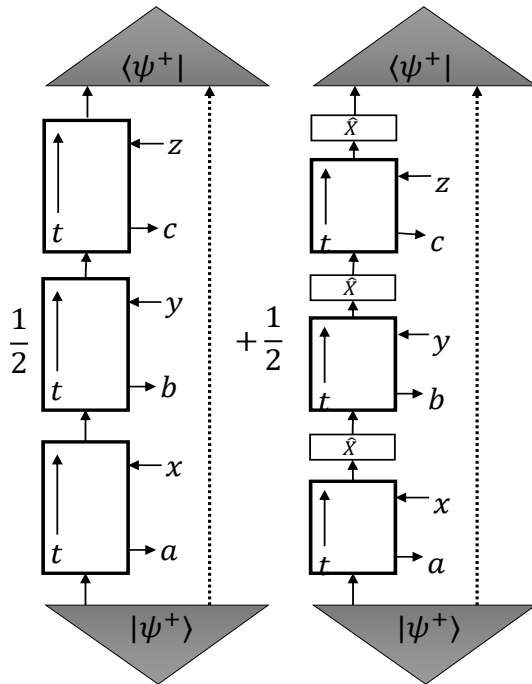
and simulate its statistical properties:

(4.64)

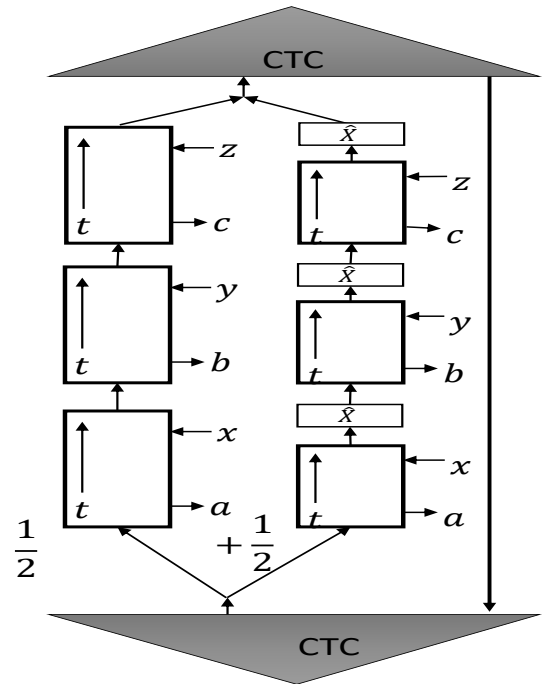
$$M(a|x)_{A_i}^{A_o} \otimes M(b|y)_{B_i}^{B_o} \otimes M(c|z)_{C_i}^{C_o} \bullet \frac{1}{2} (M_{A_o}^{B_i} \otimes M_{B_o}^{C_i} \otimes M_{C_o}^{A_i} + \bar{M}_{A_o}^{B_i} \otimes \bar{M}_{B_o}^{C_i} \otimes \bar{M}_{C_o}^{A_i}) = p(a, b, c|x, y, z)$$

so after many many runs we generate the causal inequality violating statistics of the previous section.

This is a demonstration of what we saw in section 1.3.2 of the introduction, and can be viewed as post-selection granting one the power to teleport states between Hilbert spaces irrespective of time ordering (which would *not* be possible in standard quantum theory).



(a) Realisation of the two-time state using a post-selection.



(b) Realisation of the two-time state using a CTC.

## 4.6 Discussions and Conclusions

We have given an explicit procedure that cannot be achieved with any classical causal model, but can be implemented by utilising post-selection. This procedure generates an extreme point of the NBTS polytope, and is tripartite. This is in stark conflict with the bipartite scenario, where it is known that no such non-causal correlations exist if the NBTS conditions for two-parties are imposed. In the bipartite case, the local NBTS conditions are sufficient to warrant a global distribution that is causal.

Perhaps the contrast is that for three or more parties, as mentioned earlier in the main text, it is not so clear how local laboratories are able to synchronise clocks between themselves. The possibility to have an adaptive protocol, where Alice's measurement outcome has an effect on the subsequent causal order means that the parties are not sure how to synchronise clocks unless they are clear on what is going on outside the labs.

It is interesting that this protocol is effectively 'classical' in that it does not require quantum interference or entanglement to achieve. We know from the previous chapter that all processes in 'vanilla' quantum theory have a classically causal model. Do all quantum processes (processes realisable without post-selection, and locally using quantum theory) in the NBTS scenario also have classical analogues?

The physical principle in this chapter is missing. The NBTS conditions, while operationally sensible are not enough to ensure that local distributions produce globally causal processes. What,

if any additional or new constraints are sufficient to ensure a classical causal explanation for any pre and post-selected procedure. A recent article [215] attempts to answer this question - with the 'natural desiderata' of logical consistency, linear and reversible dynamics, computational tameness, and sensible quantum-to-classical transitions all proving insufficient. This question is especially pertinent if one has the view that the universe is a post-selected state, and we are just simply at a point not mature enough to observe this fact. In some sense, we now know how quantum theory is causal, but we do not know 'why' it is causal. Such, I imagine, will remain hot topic for the philosophy of science for many years to come.





## THE RESOURCE THEORY OF POST-SELECTION

This chapter is unseen work, and was completed at the University of Bristol in conjunction with my supervisor Dr. Anthony J Short. In it, we outline a new resource theory- the resource being post-selection, which while not a quantum resource theory, does possess some of the nice qualities and features of a resource theory. This question of whether this was a valid and interesting object to investigate was my original idea, and is based in the broader topic of resource theories in quantum theory.

### 5.1 Introduction

In this chapter, we address the question of which resources can lead to causal inequality violation, among other things. We know that quantum theory contains no resources that can violate a causal inequality, and we have proven it in chapter three. Entanglement, non-locality, coherence, superposition - it does not matter - none of these resources can lead to causal inequality violation. In order to violate causal inequalities, we have to look outside of quantum theory.

The candidate resource we will study is, as one could probably guess from the theme of this thesis, is post-selection. Post-selection is enormously powerful, it has not only the power to lead to causal inequality violation, but also as we have shown can violate the second law of thermodynamics, and has complexity theoretical power equivalent to the access to a closed-time-like-curve [216]. It also has been shown to be of practical importance in cavity QED [217], noiseless amplification techniques in communication theory [218] and in novel amplification techniques used in precision measurements [219], state-of-the-art metrology protocols [220], as well as in the process of weak measurement. Some scientists argue that benefits to the weak measurement of post-selection remain unclear, such as the controversy surrounding weak values

and parameter estimation [221, 222]. What is clear, is that weak measurements represent a fundamentally quantum effect [223, 224], do not represent a simple conditional average [225] and, have a history that is intimately connected to post-selection.

There is a dark side to post-selection. One that is commonplace outside of physics, even reaching into in the media, into other sciences, into statistics and into law. To give an example that is prevalent (this fact is obvious to anybody who has ever taught an experimental physics course!) is when it so occurs that data pertaining to a particular run of an experiment is excluded from the statistics that are used to calculate results in the experiment, be it the refutation of a null hypothesis or the calculation of some physical quantity with error bars. One must be *very* careful when omitting outliers [226], and have good reason for doing so. If there are no such good reasons for doing so one must be cautious. An example from history where this goes sideways can be found in Feynman's analysis of the famous Millikan oil drop experiment [227], is that the correct value for the ratio of charge to mass of the electron was delayed by many years when, after making a blunder by using the incorrect value for a physical constant of viscosity in the original experiment, scientists of lesser repute than Millikan began fudging data of subsequent experiments to better replicate Millikan's results. How did they do this fudging - post-selection (ignoring results that were far from Millikan's and so making their data better agree with the original prestigious, but incorrect, charge-mass ratio).

Of course, looking at only a subset of outcomes (experimental post-selection) and actually deleting the record of information (fundamental post-selection) have radically differing interpretations. The latter, which is yet to be observed in nature, if it ever will, represents a violation of one of the most fundamental facts we know today [228], that information is physical and processing it costs work. The former, while less gregarious from the perspective of new physics, has wide applicability. For instance, very recent work has shown that a quantum computer constructed from only singlet or triplet type measurements and single qubit gates is universal for computation [229], under the proviso one is allowed to post-select. This is interesting since it gives a way to measure, say, the state of two atomic spins and preserve the quantum information pertaining to each spin individually, and process on the entangled subspace of the pair of particles. An essential ingredient to this method is post-selection, and this might make the dream of quantum computation a step closer to realising, since (unlike other schemes) the fundamental operations are simple to implement and protect the logical operations by symmetries of the physical Hamiltonian.

We have seen in previous chapters that such practices, spending the resource of post-selection can have fantastic consequences. One must acknowledge when and how one is doing this. The main hope of this chapter is to quantify the resource of post-selection in the language of a resource theory, so that one may better quantify when and how it is spent. We will do this in the language of two-time states.

The remainder of this chapter is structured as follows. In the next section, section 5.2 we

outline what a resource theory is, how to spot one, and some desirable qualities that they should have. We then give an example of a quantum resource theory in section 5.3. We then give the resource theory of post-selection in section 5.4 and a discussion of the results and some open questions in 5.5.

## 5.2 Resource Theories

The crux of value, is in scarcity. Gold is valuable since it is rare, and therefore hard to obtain. A PhD is valuable, since it requires one to undergo a difficult process, and the people that do it are infrequent in the population. A rifle is valuable in some situations, since it is rare to survive a war without one. Resource theories very generally describe such scenarios. Actions or objects (the distinction between these is interesting [123]) can be categorised as free or costly, and then the actions/objects outside of the class of free things obtain a value.

A concrete example of a quantum resource theory is the theory of entanglement. Permitting as free operations local operations on quantum systems, and classical communication between the parties (LOCC) entangled states become resourceful. This is because entanglement can only be generated by global operations on quantum systems. For example, imagine Alice and Bob are in separate laboratories that are connected by a telephone (a classical communication channel) but no quantum communication channel between them exists (such as a fibre optic cable). They both have a spin  $1/2$  system such as an electron in their labs. Then, entangled states such as  $|\phi^+\rangle = 1/\sqrt{2}(|00\rangle + |11\rangle)$  are resourceful. They can create correlations using (many copies of) this state that cannot be generated by using LOCC.

Inspired by the successes of entanglement theory viewed as a resource, there have been a plethora of resource theories proposed over the past few years. There are resource theories for entanglement [230], non-locality [35], coherence [231], asymmetry [232], thermodynamics [12], non-gaussianity [233] and even contextually [234]. Chitambar and Gour [123] give four reasons that resource theories have become so widespread in modern quantum information theory, which we state here;

1. **Practical.** Resource theories are enormously useful when one wishes to consider what is feasible and not feasible in the real world. They can be constructed to consider a restriction of attention to current technological capabilities, as opposed to some grand theory which may be wide in scope but fail in specificity.
2. **Comparison.** We can use resource theories to compare between different quantum systems and states and establish a pre-order of usefulness among these states. We can answer questions of not only ‘if?’ but ‘how much?’.
3. **Fine graining.** This property, allows the lens of resource theories to get at the heart of what fundamental processes drive phenomena. In the example of the phenomena being

teleportation, one can look to the resource theory of entanglement, and see instantly that the transmission of quantum states, under LOCC restriction requires entanglement.

4. Generalisation. By capturing the essence of a particular object in one resource theory, one often finds applications to resource theories in general. As an example, entanglement reversibility has connections to thermodynamic resource theories [235, 236].

We now formalise the definition of a quantum resource theory. This will lead us to the ‘golden rule’ for resource theories. We will then go outside of quantum theory to define a resource theory of post-selection. We follow [123] for the definition of a quantum resource theory.

**Definition 5.1.** *Let  $\mathcal{O}$  be a mapping that assigns to any two input and output physical systems  $A$  and  $B$  with corresponding Hilbert spaces  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , a unique set of CPTP operations  $\mathcal{O}(A \rightarrow B) := \mathcal{O}(\mathcal{H}_A \rightarrow \mathcal{H}_B) \in \mathcal{Q}(A \rightarrow B)$ .  $\mathcal{Q}$  is the set of quantum channels, which are the CPTP maps in the set of bounded linear maps from density matrices to density matrices. Let  $\mathcal{F}$  be the induced mapping  $\mathcal{F}(\mathcal{H}) := \mathcal{O}(\mathbb{C} \rightarrow \mathcal{H})$  where  $\mathcal{H}$  is an arbitrary Hilbert space. Then the tuple  $\mathcal{R} = (\mathcal{F}, \mathcal{O})$  is called a quantum resource theory if the following two conditions hold:*

1. *For any physical system  $A$ , the set  $\mathcal{O}(A) := \mathcal{O}(A \rightarrow A)$  contains the identity map.*
2. *For any three systems  $A$ ,  $B$  and  $C$ , if  $\Phi \in \mathcal{O}(A \rightarrow B)$  and  $\Lambda \in \mathcal{O}(B \rightarrow C)$ , then  $\Lambda(\Phi) \in \mathcal{O}(A \rightarrow C)$ .*

*In a quantum resource theory, the set  $\mathcal{F} \subset \mathcal{Q}$  defines the set of free states, and the elements of  $\mathcal{Q}/\mathcal{F}$  are called the resource states. Likewise, the CPTP maps in  $\mathcal{O}(A \rightarrow B)$  are called free operations and the CPTP maps not in  $\mathcal{O}(A \rightarrow B)$  are dynamical resources.*

The physical interpretation of definition (5.1) is important to establish. Quantum systems are held by one agent or distributed among agents. The quantum resource theory models what the agent or agents can accomplish given some physical restrictions, such as technological or experimental limitations (in a quantum computation this could be the gate set permitted by a device does not contain the unitary we need to implement). The operations the parties can perform for free, is described by  $\mathcal{O}(\mathcal{H}_A \rightarrow \mathcal{H}_B)$ . The first condition then says simply that doing nothing should be considered free. One finds it hard to argue against this intuition. The second condition says that  $\Lambda(\Phi(\bullet))$  is free whenever both  $\Lambda(\bullet)$  and  $\Phi(\bullet)$  are both free. Free operations can be performed in any order, and any number of times and the resultant operation should always be free. A corollary to this is the ‘golden rule’ for resource theories.

**Definition 5.2.** *The golden rule of resource theories say that ‘For any two physical systems  $A$  and  $B$ , if  $\Phi \in \mathcal{O}(A \rightarrow B)$  and  $\rho \in \mathcal{F}(A)$ , then  $\Phi(\rho) \in \mathcal{F}(B)$ ’.*

Stated colloquially, the free operations cannot convert a free state to a resourceful state. This does not in any way mean that resourceful states do not play a functional role in the theory (it is clear that resourceful operations are indeed useful from the definition or from our entanglement

example). On the contrary, given a resourceful state the agents with access to it can use it to circumvent the restrictions present in the free only operations and states of the resource theory. Given a  $\sigma \notin \mathcal{F}$ , there may exist maps  $\Phi \in \mathcal{O}(AB)$  and  $\Lambda \in \mathcal{O}(A)$  such that  $\Phi(\rho \otimes \sigma) = \Lambda(\rho)$ . This makes clear the statement of  $\sigma$  being resourceful - sigma, not being free, is facilitating a non-free operation on  $\rho$  through the action of a larger, free operation.

The above definition of a resource theory lays out what it is, and answers the ‘if?’ type questions as promised. When is a given transformation possible for example? If it is either free, or one can spend some resourceful state to implement it. When is a given state possible to generate? If it is either free, or one can spend some dynamical resource to generate it. However, the ‘how much?’ type questions are left wanting. To this end, we will now proceed with quantifying resources. This quantification will be done in terms of a resource monotone, which is a function that is zero on the free states and larger on states which have more resource and smaller on states with less of the resource.

Let us be more precise. An axiomatic approach to resource theories is the most enlightening in this regard. Central to this approach is defining a few properties that a quantum resource theory must have, and a few additional properties that are desirable. This approach centres around a resource monotone which we call  $R$ . The quantum resource theory is defined for a given Hilbert space  $\mathcal{H}$ , and as such should be able to characterise any density operator  $\rho \in \mathcal{H}$ . The function  $R$  then is a non-negative map from the space of  $\rho$  to the non-negative real numbers  $\mathbb{R}_{\geq 0}$ , giving the value of the resource for each state. In practice, one is satisfied with restriction the domain of  $R$  and focus on a class of states  $\rho$ , although this restriction should be lifted wherever possible to form the most complete and general resource theory.

The axioms satisfied by  $R$  are listed below. We distinguish [123] between the first two axioms which we deem essential and the final three, which we deem desirable. This distinction will prove useful later, when we are able to give properties for the resource theory of post-selection which are ‘almost’ desirable.

1. The first axiom is that given a resource measure  $R$ , a set of free states  $\rho \in \mathcal{F}$ , and a system  $A$ , then

$$(5.1) \quad \rho \in \mathcal{F}(A) \implies R(\rho) = 0$$

which is to say, the free states contain no resource. If this implication is two-way, i.e. the  $\implies$  becomes an  $\iff$ , then the resource theory is called faithful. Note many examples are present in the literature where this does not hold. It may be possible for a resourceful state to provide no operational advantage in a given task over a free state, such as in the theory of entanglement distillation [237], where the distillable entanglement vanishes for all bound entangled states (which are still entangled non-the-less). This, faithfulness is appealing, though not essential. The property in equation (5.1) is however essential.

2. The resource monotone should be non-increasing under free operations. This, is the defining quality of a free operation. It is also the reason we call  $R$  a monotone. A function  $R : \mathcal{H} \rightarrow \mathbb{R}_{\geq 0}$  is a valid monotone if for any  $\Phi \in \mathcal{O}(A \rightarrow B)$  and  $\rho \in \mathcal{Q}$  it holds that

$$(5.2) \quad R(\rho) \geq R(\Phi(\rho)).$$

Note that there is not a requirement that we can actually measure the monotone.

Certain resource theories appear with measurements outside of the physical operations (thermodynamics being the prime example [238]). There do exist resource theories where control are central however and are discussed at length in [49].

For quantum resource theories where two states  $\rho$  and  $\sigma$  are inter-convertible, monotonicity implies that  $R(\rho) = R(\sigma)$  whenever  $\rho$  and  $\sigma$  are free. We can then always satisfy axiom one by shifting  $R$  appropriately.

3. Convexity is often a desirable property of resource theories. This states that

$$(5.3) \quad R\left(\sum_i p_i \rho_i\right) \leq \sum_i p_i R(\rho_i).$$

Physically this states that by coin tossing and mixing resourceful states depended on the outcome, one should never be able to increase the resource. For the key work of this chapter, this property is violated. However, we are able to show a weaker form of convexity that does indeed hold.

4. Another desirable property for a quantum resource theory is sub-additivity. Physically, this states that having a system  $A$ , carrying some value of resource  $\omega$ , and another system  $B$  carrying resource  $\alpha$ , then the joint system  $A \otimes B$  should not contain more resource than  $\omega + \alpha$ . Clearly, this property while intuitive will be violated by any quantum resource theory that admits superactivation [239, 240], where the system is not totally determined by the constituent parts. This is in some ways a classical property of quantum resource theories so we dub it desirable but not essential. A formal definition of sub-additivity is that

$$(5.4) \quad R(\rho \otimes \sigma) \leq R(\rho) + R(\sigma).$$

Finally, we comment that a resource theory is additive when equality holds in the above.

5. The final desirable property that a quantum resource theory could have is asymptotic continuity. We do not touch on this property much in this work, however we state it here for completeness. Hand waving, asymptotic continuity states that if we perturb a state by a small amount, we are almost sure that the resource content will not change very much. Although, this would just be continuity, the asymptotic part means that this

convergence is calculated relative to the dimension of the system  $A$ . A function  $R$  is said to be asymptotically continuous [241] if

$$(5.5) \quad |R(\rho) - R(\sigma)| \leq K\epsilon \ln(\dim(\mathcal{H}_A)) + c(\epsilon),$$

where  $K$  is a constant to be determined,  $\rho$  and  $\sigma$  are supported by  $\mathcal{H}_A$ ,  $\epsilon = \frac{1}{2}\|\rho - \sigma\|_1$  and  $c(\epsilon)$  has the property that  $\lim_{\epsilon \rightarrow 0} c(\epsilon) = 0$  and depends only on  $\epsilon$ . (In particular, it does not depend on dimension.). One can show that the Von-Neumann entropy is asymptotically continuous, as revealed by the Fannes-Audenaert Inequality [242, 243]. Asymptotic continuity is important when one considers a many-to-many copy conversion rate between resourceful states, a task we will not discuss much in this work. However, it is noteworthy to mention that the rates of conversion in the asymptotic setting can be quite different from the rates in the single shot case, which are determined entirely by the monotonicity property.

With this characterisation in place, we will first discuss an example of a quantum resource theory to aid clarity. After this, we will be ready to step outside quantum resource theories to general resource theories, and formalise the resource theory of post-selection.

### 5.3 Quantum Resource Theory Example

In this section, we will cement our abstract discussions with a concrete example of a quantum resource theory. The example we will proceed with here is the theory of entanglement. In this theory, the physical scenario of interest is where-in spatially separated parties are permitted to freely exchange classical information - but quantum systems are held locally (and processed by CPTP maps on the subsystems). The global structure of the maps that one can implement is therefore the class of LOCC, with LOCC representing free operations in the quantum resource theory of entanglement.

An LOCC map might in general be quite complex - but this is hardly surprising, since they contain all possible local quantum physics! LOCC operations are built from adaptive protocols where a local measurement is performed followed by a global broadcast of the measurement outcome. One typically does not consider the Shannon theory in the scenario (the classical resources of the broadcast) since, with reference to the technological and practical nature of the resource theories, if not free, the classical communication can be considered very very cheap. Every LOCC operation takes the form

$$(5.6) \quad \Phi(\bullet) = \sum_k (\otimes_i M_{k,i}^{A_i}) \bullet (\otimes_i M_{k,i}^{A_i})^\dagger$$

where  $M_{k,i}^{A_i}$  is the measurement operator on party  $A_i$ , conditioned on measurement outcomes  $k$ . Worth noting is that precise analysis of this map in the networked scenario is not without interest, one can see that there will be some conditionals that cannot depend on others via causal



restrictions. This is an active topic of research [244, 245] which we leave as an open question. It is also known that LOCC is not a closed set - there exist sequences of protocols involving repeated applications of LOCC operations that converge to an operation that cannot be implemented through finite round LOCC [246].

The free states in the resource theory of entanglement are the separable states, called *SEP*. States within *SEP* always have the form

$$(5.7) \quad \rho = \sum_k p_k \rho_{1,k}^{A_1} \otimes \dots \otimes \rho_{N,k}^{A_N}$$

where  $\rho_{i,k}^{A_i}$  is an arbitrary quantum state for party  $A_i$ . It is not hard to show, through the tensor product structure that the operations (5.6) leave the structure of (5.7) invariant. Hence, we have achieved the defining quality for a resource theory and our golden rule finds itself obeyed. The resource theory also satisfied some of the other desiderata. The set *SEP*, unlike *LOCC* is indeed closed. Also, once one has appealed to specific monotones for the theory, such as entanglement measures like the entanglement of formation, the theory is convex.

Despite these nice properties, the resource theory of entanglement is not without its issues. Determining membership to *SEP* is a hard problem - NP hard in fact [247]. In investigations of this phenomena, the researchers of entanglement often resort to computationally tractable measures of ‘separability’ - in low dimensional systems the most famous being the positivity of the partial transpose condition (PPT) which states that  $\rho^{\Gamma_i} \geq 0$  for any state in *SEP*, where  $\Gamma_i$  indicates a transpose on system  $A_i$ . In higher dimensions that  $2 \otimes 3$  or  $3 \otimes 2$ , there exist states however which satisfy PPT but are still entangled. Such is the topic of active research.

## 5.4 Resource Theory of Post-Selection

What has preceded this section has all been to do with quantum resource theories. With caution we now step outside this paradigm. A quantum resource theory is so named since the set of ‘all possible’ operations and states in definition (5.1) is  $\mathcal{Q}$  - quantum states and CPTP operations. But as we have seen, post-selection permits the simulation of things, such as causal inequality violation and violations of the second law of thermodynamics, that are outside of  $\mathcal{Q}$ . Hence, our resource theory of post-selection must be able to ‘deal with’<sup>1</sup> elements of Hilbert space that are outside of  $\mathcal{Q}$ .

As mentioned before, in order to define a resource theory, we must define the free states and operations. In the resource theory of post-selection, we will assign the free states as exactly  $\mathcal{Q}$ , and the free operations as CPTP maps from  $\mathcal{Q}$  to  $\mathcal{Q}$ . The objects in the set  $\mathcal{Q}$  do not require any post-selection to implement or create so it makes sense to define them as free, quantum theory does the work for us! We will proceed more formally, in the two-time language of the

---

<sup>1</sup>Taken to mean, able to be processed by and understood within the native language of the resource theory. This is not to be confused with the way quantum theory ‘deals with’ non-normalised plane waves for example.

previous section and keep the conventions surrounding the labelling of Hilbert spaces and the dot operation.

The free states within our theory are all well normalised, or at least normalise-able;

$$(5.8) \quad \mathbb{1}_A \bullet \eta^B \propto 1 \iff \eta \in \mathcal{Q},$$

which holds for quantum states. The free operations within our theory all carry the future identity preserving property as their defining quality,

$$(5.9) \quad \mathbb{1}_B \bullet \eta_A^B \propto \mathbb{1}_A \iff \eta \in \mathcal{Q},$$

which holds for all quantum channels.

We should take a brief moment to address normalisation within the two-time formalism. Physically, normalisation in this theory is irrelevant. What matters are the ratio's between probabilities, not their absolute magnitude [55]. Indeed, correct normalisation of two-time states does not automatically imply correct normalisation of the probabilities of measurement outcomes. The overall probability to prepare two-time states depends upon the success of the post-selections involved in doing so, and these probabilities in standard quantum theory depend themselves upon the measurements that are made in the intermediate times. In standard quantum theory to contract, the probability to prepare a quantum state (for example, for later use in a quantum algorithm) is equal to 1 (even though in practice this is hardly ever the case!). The two time formalism averts this problem since it correctly calculates relative probabilities (this event has twice the chance of happening of this other, complimentary event) and only afterwards normalises the probabilities (here we would assign  $p = 2/3$  to the first event and  $p = 1/3$  for the second). We stress this difference by defining  $\mathcal{Q}$  to be the set of any quantum objects that appear with any normalisation constant, and reserve the notation  $\tilde{\mathcal{Q}}$  for the states which are normalised *ex pre facto*. Hence, for  $q, p \in \tilde{\mathcal{Q}}$ ,  $q \bullet p = 1$ , whereas for  $q, p \in \mathcal{Q}$ ,  $q \bullet p = c$  where  $c$  is some real number. We also adopt the convention that any contraction in the dot without system labels is total.

**Definition 5.3.** *The total contraction implicitly denoted without system labels by  $p(q) = q \bullet p$  is taken to mean that  $p$  provides a contraction over all hanging inputs and outputs of  $q$ . As an example, if  $q = q_{A,B}^{C,D} \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}^C \otimes \mathcal{H}^D$ , then implicitly  $p = p_{C,D}^{A,B} \in \mathcal{H}^A \otimes \mathcal{H}^B \otimes \mathcal{H}_C \otimes \mathcal{H}_D$ . Note that we adopt the convention of the last chapter, where the Roman Hilbert space  $\mathcal{H}^A$  is related to the calligraphic Hilbert space  $\mathcal{H}^{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{A}}$ . Total contractions are scalars.*

We are now ready to introduce our resource monotone.

**Definition 5.4.** *Our resource monotone can be quantified with the following optimisation problem,*

$$(5.10) \quad R(M) = \min_{\tilde{M} | (M + \tilde{M}) \in \mathcal{Q}} \max_{q \in \mathcal{Q}} \frac{q \bullet \tilde{M}}{q \bullet (M + \tilde{M})}.$$

Where  $M$  is any pre-or-post-selected channel (described by a CP, but not necessarily TP map) from  $\mathcal{H}_A \otimes \mathcal{H}^{A^\dagger}$  to  $\mathcal{H}_B \otimes \mathcal{H}^{B^\dagger}$ ,  $\tilde{M}$  is a complimentary (and in further works referred to as the completion

of  $M$ ), potentially also post-selected, CP map from  $\mathcal{H}_A \otimes \mathcal{H}^{A^\dagger}$  to  $\mathcal{H}_B \otimes \mathcal{H}^{B^\dagger}$  that completes the sum  $M + \tilde{M}$  to a physically implementable, non post-selected quantum process. This implies that all hanging indices of  $M$  and  $\tilde{M}$  are identical. The state  $q$  ranges over all quantum states achieving a total contraction and settles on the specific state which makes the post-selection most likely to succeed.

First we note that we immediately see that the free states do indeed have zero resource as desired, since

$$(5.11) \quad R(q \in \mathcal{Q}) = \min_{\tilde{q} | (q+\tilde{q}) \in \mathcal{Q}} \max_{q \in \mathcal{Q}} \frac{q \bullet \tilde{q}}{q \bullet (q + \tilde{q})},$$

we are free to pick  $\tilde{q}$  to be the zero state, which reduces the above min-max problem to 0. Hence,  $R(q \in \mathcal{Q}) = 0$ . Before we discuss how this resource monotone forms an appropriate characterisation of a resource theory, we will give a short example that I hope will be illustrative.

Consider the CPTP map  $M$ , described by the Kraus operator description,

$$(5.12) \quad M_A^B = K_{0\mathcal{B}^\dagger}^{\dagger\mathcal{A}^\dagger} \otimes K_{0\mathcal{A}}^{\mathcal{B}} + K_{1\mathcal{B}^\dagger}^{\dagger\mathcal{A}^\dagger} \otimes K_{1\mathcal{A}}^{\mathcal{B}} \quad \text{where } K_{0\mathcal{A}}^{\mathcal{B}} = |0\rangle\langle 0| \\ K_{1\mathcal{A}}^{\mathcal{B}} = |1\rangle\langle 1|,$$

which could describe a measurement in the computational basis. Now consider the post-selected CP map  $\hat{M}$ , described by the Kraus operators

$$(5.13) \quad \hat{M}_A^B = \hat{K}_{0\mathcal{B}^\dagger}^{\dagger\mathcal{A}^\dagger} \otimes \hat{K}_{0\mathcal{A}}^{\mathcal{B}} + \hat{K}_{1\mathcal{B}^\dagger}^{\dagger\mathcal{A}^\dagger} \otimes \hat{K}_{1\mathcal{A}}^{\mathcal{B}} \quad \text{where } \hat{K}_{0\mathcal{A}}^{\mathcal{B}} = |0\rangle\langle 0| \\ \hat{K}_{1\mathcal{A}}^{\mathcal{B}} = \sqrt{\alpha} |1\rangle\langle 1| \quad \text{where } 0 < \alpha < 1.$$

We can build<sup>2</sup> the map CP(5.13) from CPTP (5.12) in the following manner. We implement the CPTP operation (5.12) by measuring a qubit in the computational basis, and record statistics. We then discard from the subset of statistics associated to observing the 1 outcome with probability  $\alpha$ . The resultant statistics are exactly those which correspond to an implementation of the map (5.13). Therefore this entire process is equivalent to one that can be described by the Kraus operators for the map  $\hat{M}_A^B$ .

It is again in general possible to build the CPTP map (5.12) exclusively from the map CP (5.13). However, we will once again have to post-select. I will leave this as an exercise for the reader - to be conducted especially after the following section upon rescaling by the identity operation. However, there is an alternative method, which will appeal to the intuition of our resource monotone.

We discuss what we mean by completion. By defining the *completion* of  $\hat{M}_A^B$  to be  $\tilde{M}_A^B = (1 - \alpha)|1\rangle\langle 1| \otimes |1\rangle\langle 1|$ , we note immediately that it would be possible to implement the map  $M_A^B$  in a simple fashion, by using the maps  $\hat{M}_A^B$  and  $\tilde{M}_A^B$ .

<sup>2</sup>simulate the application of

It is easy to verify that this aggregation implements  $M_A^B$ , explicitly since

$$(5.14) \quad \hat{M}_A^B + \tilde{M}_A^B = \sum_i \hat{K}_{i,\mathcal{A}}^{\dagger\mathcal{B}} \otimes \hat{K}_{i,\mathcal{A}}^{\dagger\mathcal{B}} + \tilde{M}_A^B = |0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1| = M_A^B.$$

This is in fact precisely what we mean when we say that  $\tilde{M}$  is a completion. The summation  $\hat{M} + \tilde{M}$  is in  $\mathcal{Q}$ . It is in fact the best completion - where best, in this context means that it is the smallest with respect to some norm, which we will introduce shortly. If the fact that this is the best completion is given, one can calculate the resource for the map  $\tilde{M}$ , to be  $R(\tilde{M}) = 1 - \alpha$ , hinting towards the operational characterisation of the resourcefulness of a post-selection being the maximum proportion of outcomes that are discarded in the set of statistics under a post-selection.

We also note some other properties of the resource measure (5.10). Firstly, it is bounded between 0 and 1. This can be seen, either through its association to a probability of discarding statistics, or through the fact that  $\tilde{M}$  is a CP map. Thus the numerator in equation (5.10) is always smaller than or equal to the denominator. As a measure of a probability, this property is not surprising, having observed this we may now claim that the resource is the best-case success probability of post-selection succeeding for a given process  $M$ .

**Definition 5.5.** *We will define a “norm” on the CP maps as*

$$(5.15) \quad \|M\| = \min_{\tilde{M}|M+\tilde{M}\in\mathcal{Q}} \bar{q} \bullet (M + \tilde{M}).$$

where  $\bar{q}$  is any quantum operation in  $\tilde{\mathcal{Q}}$  which makes the contraction total (since  $M + \tilde{M} \in \mathcal{Q}$  and the quantum set is closed under composition this ensures that our “norm” provides a scalar). We have also used norm in inverted commas, since strictly norms are typically on vector spaces, not sets.

The definition (5.5) can be shown to obey the triangle inequality since

$$(5.16) \quad \begin{aligned} \|M_1 + M_2\| &= \min_{\tilde{M}|M_1+M_2+\tilde{M}\in\mathcal{Q}} \bar{q} \bullet (M_1 + M_2 + \tilde{M}) \\ &\leq \min_{\tilde{M}_1, \tilde{M}_2 | M_1+M_2+\tilde{M}_1+\tilde{M}_2 \in \mathcal{Q}} \bar{q} \bullet (M_1 + M_2 + \tilde{M}_1 + \tilde{M}_2) \\ &\leq \min_{\tilde{M}_1 | M_1+\tilde{M}_1 \in \mathcal{Q}} \bar{q} \bullet (M_1 + \tilde{M}_1) + \min_{\tilde{M}_2 | M_2+\tilde{M}_2 \in \mathcal{Q}} \bar{q} \bullet (M_2 + \tilde{M}_2) \\ &= \|M_1\| + \|M_2\|. \end{aligned}$$

and positivity since the map  $M$  is CP. Furthermore, we are able to verify that  $\|cM\| = c\|M\|$  for all positive constants  $c$ , since the map  $M$  itself is CP. We note that quantum object has unit norm,

$$(5.17) \quad \|M \in \tilde{\mathcal{Q}}\| = \min_{\tilde{M}|M+\tilde{M}\in\mathcal{Q}} \bar{q} \bullet (M + \tilde{M}) = \bar{q} \bullet M = 1.$$

since, by definition, normalised quantum operations and states satisfy the relation  $\bar{q} \bullet \tilde{M} = 1$ . With this definition (5.15), we note that the resource is equal to

$$(5.18) \quad R(M) = 1 - S(M) = 1 - \min_{\bar{q} \in \tilde{\mathcal{Q}}} \frac{\bar{q} \bullet M}{\|M\|}.$$

We will occasionally refer to the quantity  $S(M) = \min_{\tilde{q} \in \mathcal{Q}} \frac{\tilde{q} \cdot M}{\|\tilde{M}\|}$  as the success probability for brevity.

We will now show that the resource (5.10) is non-increasing under a class of free operations. The operations that are free within this theory take two forms. The first are the quantum operations. These can be stated heuristically as either the map gets a quantum circuit appended to the input or output as a post or pre-processing. We will denote this map with  $T$ , i.e. the free transformations are defined as  $T(M_A^B) = Q_C^A \cdot M_A^B$  or  $T(M_A^B) = Q_B^C \cdot M_A^B$  for  $Q \in \mathcal{Q}$ . The second type of free operations, are uniform discarding of statistics. These maps, which we denote with  $P$  are always proportional to the identity over the output of the operand map. Physically, we can justify that these operations should be free since they do not require a measurement of the quantum system in order to implement their post-selection. One can do it purely with access to classical randomness, and tossing coins to pre-determine if a given measurement outcome will or will not be determined (these operations are called trivial post-selections [209, 198]). Unlike the case of a true post-selection, where the measurement results are either kept or discarded according to some measurement of a quantum system that *depends on the system*, the maps  $P$  are always system invariant and as such, we permit them to be free. Therefore, we can always represent  $P(M_A^B)$  by  $P(M_A^B) = cM_A^B$ , where  $c$  is a constant.

We can now show that the resource is non-increasing under these operations. Let  $M^*$  be the (not necessarily unique) completion which minimises equation (5.10). The resource is then given by

$$\begin{aligned}
 R(M) &= \max_{q \in \mathcal{Q}} \frac{q \cdot M^*}{q \cdot (M + M^*)} \\
 &\geq \max_{p \in \mathcal{Q}} \frac{p \cdot Q \cdot M^*}{p \cdot Q \cdot (M + M^*)} \\
 &= \max_{p \in \mathcal{Q}} \frac{p \cdot (Q \cdot M^*)}{p \cdot (Q \cdot M + Q \cdot M^*)} \\
 &\geq \min_{\tilde{M} | (Q \cdot M + \tilde{M}) \in \mathcal{Q}} \max_{p \in \mathcal{Q}} \frac{p \cdot \tilde{M}}{p \cdot (Q \cdot M + \tilde{M})} \\
 (5.19) \quad &= R(T(M)).
 \end{aligned}$$

Here,  $T(M) = Q \cdot M$ , and we have used that since  $\mathcal{Q}$  is closed under composition, there exists  $q = p \cdot Q$  for all  $p, q$  in the set. Therefore, the resource monotone (5.10) is indeed a good monotone for the  $T$  type free operations.

This result means that, since the resource is strictly bounded between 0 and 1, appending a quantum circuit to a quantum circuit will not increase the value of the resource (beyond its minimal value of 0). Hence, constructing a quantum circuit from a set of primitive elements is not a task which one needs to ‘spend any’ post-selection to implement.

This is telling us fundamental post-selections are not characterised by quantum theory, since all quantum objects do not carry any value under  $R$ . One is inclined to say that post-selection has a unique character as being among physical operations one can simply conceive, yet not describe

within quantum theory (and one must consider the role of the observer if one wishes to describe an experimental post-selection). Post-selection is certainly one of the most distinctive sources of paradoxes that one can tackle quantum theory with, and has the character of resistant to interpretation without considering the agency by which the post selection is performed. All that can be said is, that once agency is properly considered, post-selection bears fruits [248].

Returning to the golden rule, we are able to show that the second type of free transformation,  $P(M)$ , saturates these bounds. In more detail,

$$\begin{aligned}
 R(P(M)) = R(cM) &= \min_{\tilde{M} | (cM + \tilde{M}) \in \mathcal{Q}} \max_{q \in \mathcal{Q}} \frac{q \cdot \tilde{M}}{q \cdot (cM + \tilde{M})} \\
 &= \min_{\tilde{M} | (cM + \tilde{M}) \in \mathcal{Q}} \max_{p \in \mathcal{Q}} \frac{\frac{1}{c} p \cdot \tilde{M}}{\frac{1}{c} p \cdot (cM + \tilde{M})} \\
 &= \min_{\frac{\tilde{M}}{c} | (M + \frac{\tilde{M}}{c}) \in \mathcal{Q}} \max_{p \in \mathcal{Q}} \frac{p \cdot \frac{\tilde{M}}{c}}{p \cdot (M + \frac{\tilde{M}}{c})} \\
 (5.20) \qquad &= \min_{\hat{M} | (M + \hat{M}) \in \mathcal{Q}} \max_{p \in \mathcal{Q}} \frac{p \cdot \hat{M}}{p \cdot (M + \hat{M})} = R(M).
 \end{aligned}$$

In other words, we do not have to spend resource to uniformly discard statistics at random. We only need to spend resource when the post-selection is in some way state dependant (such as in a conditioning upon measurement outcomes). The resource monotone (5.10) is a good monotone for the  $P$  type free operations as well. Therefore, the resource monotone (5.10) is indeed a good monotone for the resource theory of post-selection.

We have now satisfied the two essential properties of a resource theory - that the free states carry no resource and that the resource is non-increasing under free operations. The golden rule of resource theories also now holds - neither

1. discarding at random
2. a quantum pre and-post processing

increase the resource value of the maps. The resource theory of post-selection, while not being a quantum resource theory (we can in fact define an almost equivalent classical theory by restricting to diagonal states and stochastic maps that are sub-normalised) is formed. It interestingly, and thankfully for this work, admits some of the desirable properties of a quantum resource theory that were outlined in the previous section.

Firstly, and for instance, we can show that although not a manifestly convex theory, a weaker form of convexity in the resource theory of post-selection is present.

**Result 5.1.** *The resource theory of post-selection is a convex theory, if one appropriately normalises the resource.*

$$(5.21) \quad \frac{\|M_1\|}{\|M_1\| + \|M_2\|} R(\bar{M}_1) + \frac{\|M_2\|}{\|M_1\| + \|M_2\|} R(\bar{M}_2) \geq R\left(\frac{\|M_1\|}{\|M_1\| + \|M_2\|} \bar{M}_1 + \frac{\|M_2\|}{\|M_1\| + \|M_2\|} \bar{M}_2\right),$$

which is a weaker form of convexity where we must be cautious to re-scale our maps by their size according to the norm (5.15),  $\bar{M} = \frac{M}{\|M\|}$ .

This gives us a way to compare mixtures of post-selected maps. We can show that under some additional considerations, it is not possible to increase the resource value for a given post-selected map through mixtures thereof. Proceeding with the proof, and appealing to the form of the resource given in (5.18) we find,

$$\begin{aligned}
 \frac{\|M_1\|}{\|M_1\| + \|M_2\|} R(\bar{M}_1) + \frac{\|M_2\|}{\|M_1\| + \|M_2\|} R(\bar{M}_2) &= \frac{\|M_1\|}{\|M_1\| + \|M_2\|} - \frac{\|M_1\|}{\|M_1\| + \|M_2\|} \min_{\bar{q}_1 \in \mathcal{Q}} \frac{\bar{q}_1 \cdot \bar{M}_1}{\|\bar{M}_1\|} \\
 &+ \frac{\|M_2\|}{\|M_1\| + \|M_2\|} - \frac{\|M_2\|}{\|M_1\| + \|M_2\|} \min_{\bar{q}_2 \in \mathcal{Q}} \frac{\bar{q}_2 \cdot \bar{M}_2}{\|\bar{M}_2\|} \\
 &= 1 - \frac{\|M_1\| \min_{\bar{q}_1 \in \mathcal{Q}} \bar{q}_1 \cdot \bar{M}_1 + \|M_2\| \min_{\bar{q}_2 \in \mathcal{Q}} \bar{q}_2 \cdot \bar{M}_2}{\|M_1\| + \|M_2\|} \\
 &\geq 1 - \frac{\min_{\bar{q} \in \mathcal{Q}} \bar{q} \cdot (\|M_1\| \bar{M}_1 + \|M_2\| \bar{M}_2)}{\|M_1\| + \|M_2\|} \\
 &\geq 1 - \frac{\min_{\bar{q} \in \mathcal{Q}} \bar{q} \cdot (\|M_1\| \bar{M}_1 + \|M_2\| \bar{M}_2)}{\|M_1 + M_2\|} \\
 (5.22) \qquad \qquad \qquad &= R(\|M_1\| \bar{M}_1 + \|M_2\| \bar{M}_2)
 \end{aligned}$$

which is equal to  $R(\frac{\|M_1\|}{\|M_1\| + \|M_2\|} \bar{M}_1 + \frac{\|M_2\|}{\|M_1\| + \|M_2\|} \bar{M}_2)$  since we have shown that rescaling by a constant does not change the value for the resource, and indeed  $\|M_1\| + \|M_2\|$  is just some (positive) constant.

Another of the desirable properties that we admit is a version of additivity. As mentioned, this property means that having access to two resourceful states should not have more resource than the sum of the states, taken individually. We will have to modify this intuition in the case of post-selection, but something very similar we will see to hold.

Taking a step back, we would expect that having access to a side channel which does nothing should not produce a more resourceful post-selection than the post-selected channel alone. Mathematically, this can be represented by the following result.

**Result 5.2.** *The value of the resource is identical if we allow a side channel which does nothing. In other words,*

$$(5.23) \qquad \qquad \qquad R(\mathbb{1} \otimes M) = R(M).$$

Indeed, this condition can be shown to hold. Consider

$$(5.24) \qquad \qquad \qquad R(\mathbb{1} \otimes M) = \min_{\tilde{M}(\mathbb{1} \otimes M + \tilde{M}) \in \mathcal{Q}} \max_{q \in \mathcal{Q}} \frac{q \cdot \tilde{M}}{q \cdot (\mathbb{1} \otimes M + \tilde{M})}.$$

It is simple to verify that by picking  $\tilde{M} = \mathbb{1} \otimes M^*$  and  $q = \mathbb{1} \otimes q^*$ , where  $M^*$  and  $q^*$  are the solutions to the min-max problem in (5.10) we recover the corresponding solution for  $R(\mathbb{1} \otimes M)$ . Hence,  $R(\mathbb{1} \otimes M) = R(M)$ .

**Result 5.3.** Analogously, having access to any quantum side channel should not make a post-selected channel more resourceful. This can be captured by the condition

$$(5.25) \quad R(q \otimes M) \leq R(M).$$

for all  $q \in \mathcal{Q}$ .

We will show how to prove this.

**Result 5.4.** The first thing to note, is that the norm (5.15) itself is sub-multiplicative over the tensor product;

$$(5.26) \quad \|M_1 \otimes M_2\| \leq \|M_1\| \|M_2\|.$$

The proof is simple. First, expanding out the definition,

$$(5.27) \quad \|M_1 \otimes M_2\| = \min_{\tilde{M} | M_1 \otimes M_2 + \tilde{M}_{12} \in \mathcal{Q}} \bar{q} \bullet (M_1 \otimes M_2 + \tilde{M}_{12})$$

Letting  $\tilde{M}_{12}^*$  be a minimiser,

$$(5.28) \quad \bar{q} \bullet (M_1 \otimes M_2 + \tilde{M}_{12}^*) \leq \bar{q} \bullet (M_1 \otimes M_2 + \hat{M}_{12})$$

where

$$(5.29) \quad \hat{M}_{12} = M_1 \otimes M_2^* + M_1^* \otimes M_2 + M_1^* \otimes M_2^*$$

and where  $M_1^*$  and  $M_2^*$  are the minimisers for the individual norm problems. Hence,

$$(5.30) \quad \|M_1 \otimes M_2\| \leq \bar{q} \bullet (M_1 + M_1^*) \otimes (M_2 + M_2^*) = \|M_1\| \|M_2\|.$$

which completes the proof of statement (5.26). We are able to use result 5.4 to show that the resource theory is sub-additive over the tensor product in the following manner. Consider the success probability

$$(5.31) \quad S(M_1 \otimes M_2) = \min_{\bar{q} \in \mathcal{Q}} \frac{\bar{q} \bullet M_1 \otimes M_2}{\|M_1 \otimes M_2\|}$$

Which, using result 5.4 can be shown to be less than

$$(5.32) \quad \begin{aligned} S(M_1 \otimes M_2) &\geq \min_{\bar{q} \in \mathcal{Q}} \frac{\bar{q} \bullet M_1 \otimes M_2}{\|M_1\| \|M_2\|} \\ &\geq \min_{\bar{q}' \in \mathcal{Q}} \frac{\bar{q}' \bullet M_1}{\|M_1\|} \min_{\bar{q}'' \in \mathcal{Q}} \frac{\bar{q}'' \bullet M_2}{\|M_2\|} = S(M_1) S(M_2) \end{aligned}$$

Hence

$$(5.33) \quad 1 - R(M_1 \otimes M_2) \geq (1 - R(M_1))(1 - R(M_2)) = 1 - R(M_1) - R(M_2) + R(M_1)R(M_2)$$



and so

$$(5.34) \quad R(M_1 \otimes M_2) \leq R(M_1) + R(M_2)$$

Which shows the important relation; that the resource theory is sub-additive over the tensor product. What is this relation (5.33) telling us physically? Consider a pair of post-selections performed locally by two parties, Alice and Bob. They run some experiment, and discard experimental runs according to a post-selection procedure (which can be different for each of them, and depended on some state which they hold locally). Now, they come together and compare what they did. One will see that a global post-selection has only one way to succeed - which is when Alice and Bob both succeed. On the other had, the post-selection has three ways to fail. Either Alice fails independently, Bob fails independently, or they both fail independently. This is what the relation (5.33) captures. If Alice locally has an  $\alpha$  chance for failing a post-selection, and Bob has a  $\beta$  chance of failing a post-selection, then there is at least a  $(1 - \alpha)(1 - \beta)$  chance that they together pass the post-selection. To reiterate, Alice can fail independently, Bob can fail independently or they can fail together. Using this, we can also validate our intuition about quantum side channels in 5.25. If  $q$  is in  $\mathcal{Q}$ , so that  $R(q) = 0$ , then  $1 - R(q \otimes M) = (1 - R(q))(1 - R(M)) = 1 - R(M)$  so that  $R(q \otimes M) \leq R(M)$  as claimed.

In order to better illustrate the previous results, let us walk through a simple example for bipartite channels. Consider the channels well described by the Kraus operators

$$(5.35) \quad \begin{aligned} K_0 &= |0\rangle\langle 0| \\ K_1 &= \sqrt{\alpha}|1\rangle\langle 1| \text{ Where } 0 < \alpha < 1 \end{aligned}$$

and a second channel described by the set of Kraus operators

$$(5.36) \quad \begin{aligned} L_0 &= \sqrt{\beta}|0\rangle\langle 0| \text{ Where } \gamma < \beta < 1 \\ L_1 &= \sqrt{\gamma}|1\rangle\langle 1| \text{ Where } 0 < \gamma < \beta \end{aligned}$$

where we have suppressed system labels, and are to understand that the channels are linear mappings from the Hilbert space  $\mathcal{H}_A \otimes \mathcal{H}^{A^\dagger}$  to  $\mathcal{H}_B \otimes \mathcal{H}^{B^\dagger}$ . We have as before,

$$(5.37) \quad M_1 = \sum_i K_i^\dagger \otimes K_i = |0\rangle\langle 0| \otimes |0\rangle\langle 0| + \alpha |1\rangle\langle 1| \otimes |1\rangle\langle 1|,$$

and

$$(5.38) \quad M_2 = \sum_i L_i^\dagger \otimes L_i = \beta |0\rangle\langle 0| \otimes |0\rangle\langle 0| + \gamma |1\rangle\langle 1| \otimes |1\rangle\langle 1|.$$

We have calculated  $R(M_1)$  previously, and found it to be equal to  $(1 - \alpha)$ . A straightforward and similar calculation for  $M_2$  shows that,

$$(5.39) \quad R(M_2) = \min_{\tilde{M}_2 | M_2 + \tilde{M}_2 \in \mathcal{Q}} \max_{q_1 \in \mathcal{Q}} \frac{q \bullet \tilde{M}_2}{q_1 \bullet (M_2 + \tilde{M}_2)} = \max_{q_1 \in \mathcal{Q}} \frac{q \bullet (\beta - \gamma) |1\rangle\langle 1| \otimes |1\rangle\langle 1|}{q \bullet \beta \mathbb{1}} = \frac{\beta - \gamma}{\beta}.$$

The min-max problem can be solved by inspection, and picking  $\tilde{M}_2 = (\beta - \gamma)|1\rangle\langle 1|$  as the completion for  $M_2$ . The norms for both of the channels can be calculated using definition (5.15),

$$(5.40) \quad \|M_1\| = 1 \text{ and } \|M_2\| = \beta.$$

Let us verify our results concerning the convexity and tensor product structure of the resource of post-selection. We start by noting that

$$(5.41) \quad \frac{\|M_1\|}{\|M_1\| + \|M_2\|} R(\tilde{M}_1) + \frac{\|M_2\|}{\|M_1\| + \|M_2\|} R(\tilde{M}_2) = \frac{1}{1 + \beta} (1 - \alpha) + \frac{\gamma}{1 + \beta} \frac{\beta - \gamma}{\beta} = 1 - \frac{\alpha + \gamma}{1 + \beta},$$

which will be useful later. We note that

$$(5.42) \quad M_1 + M_2 = (1 + \beta)|0\rangle\langle 0| \otimes |0\rangle\langle 0| + (\gamma + \alpha)|1\rangle\langle 1| \otimes |1\rangle\langle 1| + \gamma|1\rangle\langle 1| \otimes |0\rangle\langle 0| + \beta|0\rangle\langle 0| \otimes |1\rangle\langle 1|$$

picking  $\tilde{M}$  to be in definition (5.10),

$$(5.43) \quad \tilde{M} = ((1 + \beta) - (\gamma + \alpha))|1\rangle\langle 1| \otimes |1\rangle\langle 1| + \gamma|1\rangle\langle 1| \otimes |0\rangle\langle 0| + \beta|0\rangle\langle 0| \otimes |1\rangle\langle 1|$$

we find

$$(5.44) \quad R(M_1 + M_2) = 1 - \frac{\alpha + \gamma}{1 + \beta}$$

so we conclude these channels are a case where the convexity bound is saturated. Next, we expect that  $(1 - R(M_1 \otimes M_2)) \geq (1 - R(M_1))(1 - R(M_2))$ . Let us verify this with a direct calculation. Since  $M_1 \otimes M_2$  is given by,

$$(5.45) \quad \begin{aligned} M_1 \otimes M_2 = & \beta|0\rangle\langle 0| \otimes |0\rangle\langle 0| + \beta\alpha|1\rangle\langle 1| \otimes |0\rangle\langle 0| + \beta\alpha|0\rangle\langle 0| \otimes |1\rangle\langle 1| \\ & + \gamma|0\rangle\langle 0| \otimes |1\rangle\langle 1| + \gamma\alpha|1\rangle\langle 1| \otimes |1\rangle\langle 1| + \gamma\alpha|0\rangle\langle 0| \otimes |1\rangle\langle 1| + \beta|0\rangle\langle 0| \otimes |1\rangle\langle 1| \end{aligned}$$

which clearly has  $\beta$  as a largest eigenvalue, we can conclude that  $\|M_1 \otimes M_2\| = \beta$ , by picking  $\tilde{M}_{12} = (\beta - \beta\alpha)|1\rangle\langle 1| \otimes |1\rangle\langle 1| + \beta\alpha|1\rangle\langle 1| \otimes |0\rangle\langle 0| + \beta\alpha|0\rangle\langle 0| \otimes |1\rangle\langle 1| + (\beta - \gamma)|0\rangle\langle 0| \otimes |0\rangle\langle 0| + (\beta - \gamma\alpha)|1\rangle\langle 1| \otimes |1\rangle\langle 1| + \gamma|0\rangle\langle 0| \otimes |1\rangle\langle 1| + \gamma\alpha|1\rangle\langle 1| \otimes |1\rangle\langle 1| + \gamma\alpha|0\rangle\langle 0| \otimes |1\rangle\langle 1|$  as the completion ( $\beta - \gamma\alpha > 0$  since  $\gamma < \beta$  and  $\alpha < 1$ ). Now, appealing to the second definition of the resource, we find that

$$(5.46) \quad \begin{aligned} R(M_1 \otimes M_2) &= 1 - \min_{\tilde{q} \in \mathcal{Q}} \frac{M_1 \otimes M_2 \cdot \tilde{q}}{\|M_1 \otimes M_2\|} \\ &= 1 - \frac{\alpha\gamma}{\beta} \end{aligned}$$

by picking  $\tilde{q} = |1\rangle\langle 1| \otimes |1\rangle\langle 1| + |0\rangle\langle 0| \otimes |1\rangle\langle 1| + |1\rangle\langle 1| \otimes |0\rangle\langle 0| + |0\rangle\langle 0| \otimes |0\rangle\langle 0|$ . Hence, as claimed  $(1 - R(M_1))(1 - R(M_2)) \geq (1 - R(M_1 \otimes M_2))$ .

The equality in this result is no accident, and is a result of using channels that describe complete basis measurements as our example. This is a general feature of the resource theory of post-selection for complete basis measurements, which we shall now prove.

**Result 5.5.** *The norm is not only sub-additive, it is additive,*

$$(5.47) \quad \|M_1 \otimes M_2\| = \|M_1\| \|M_2\|.$$

whenever  $M_1$  and  $M_2$  have the forms

$$(5.48) \quad \begin{aligned} M_{1A}^B &= \sum_{i=1}^n c_i X_{iA}^B \\ M_{2C}^D &= \sum_{i=1}^m d_i Y_{iC}^D \end{aligned}$$

where  $X_i, Y_i$  are operators that sum to an element of  $\mathcal{Q}$ , i.e.  $\sum_i X_{iA}^B \bullet \mathbb{1}_B = \mathbb{1}_A$ ,  $\sum_i Y_{iC}^D \bullet \mathbb{1}_D = \mathbb{1}_C$  or equivalently that  $\sum_i X_{iA}^B = q$ ,  $\sum_i Y_{iC}^D = p$  for  $q$  and  $p \in \mathcal{Q}$ .

To prove this, consider now, with system labels included for clarity (all contractions are still total, this is just notation):

$$(5.49) \quad \|M_{1A}^B \otimes M_{2C}^D\| = \min_{\tilde{M} | M + \tilde{M} \in \mathcal{Q}} \bar{q} \bullet (M_{1A}^B \otimes M_{2C}^D + \tilde{M}_{12AC}^{BD}).$$

Let us express  $M_1$  and  $M_2$  as

$$(5.50) \quad \begin{aligned} M_{1A}^B &= \sum_{i=1}^n c_i X_{iA}^B \\ M_{2C}^D &= \sum_{i=1}^m d_i Y_{iC}^D \end{aligned}$$

where  $c_1 \geq c_2 \geq \dots \geq c_n$ ,  $d_1 \geq d_2 \geq \dots \geq d_m$  and  $c_i, d_i \geq 0$ . We also have the relations that the  $\sum_i X_i = \sum_i Y_i \in \mathcal{Q}$ . Then, a direct calculation shows that

$$(5.51) \quad \begin{aligned} \|M_1\| &= c_1 \\ \|M_2\| &= d_1. \end{aligned}$$

Calculating the norm for the tensor product  $M_{1A}^B \otimes M_{2C}^D$  on the other hand leads to

$$(5.52) \quad \begin{aligned} \|M_{1A}^B \otimes M_{2C}^D\| &= \min_{\tilde{M} | M_1 \otimes M_2 + \tilde{M}_{12} \in \mathcal{Q}} \bar{q} \bullet (M_1 \otimes M_2 + \tilde{M}_{12}) \\ &= \min_{\tilde{M} | M_1 \otimes M_2 + \tilde{M}_{12} \in \mathcal{Q}} \bar{q} \bullet \left( \sum_{i=1}^n c_i X_{iA}^B \otimes \sum_{j=1}^m d_j Y_{jC}^D + \tilde{M}_{12} \right) \\ &= \min_{\tilde{M} | M_1 \otimes M_2 + \tilde{M}_{12} \in \mathcal{Q}} \bar{q} \bullet \left( \sum_{i=1}^n \sum_{j=1}^m c_i d_j X_{iA}^B \otimes Y_{jC}^D + \tilde{M}_{12} \right). \end{aligned}$$

Clearly the smallest possible  $\tilde{M}_{12}$  that completes our tensor product of maps to something in the quantum set, recalling that it must be a positive operator, is found by direct enumeration of all possible completion operators constructed from the set  $\{X_{iA}^B \otimes Y_{jC}^D\}$  to be

$$(5.53) \quad \tilde{M}_{12} = \sum_{i,j \neq 1,1}^{n,m} (c_i d_1 - c_i d_j) X_{iA}^B \otimes Y_{jC}^D.$$

Thus

$$\begin{aligned}
 \|M_{1A}^B \otimes M_{2C}^D\| &= \bar{q} \cdot \left( \sum_{i=1}^n \sum_{j=1}^m c_i d_j X_{iA}^B \otimes Y_{jC}^D + \sum_{i,j \neq 1,1}^{n,m} (c_i d_1 - c_i d_j) X_{iA}^B \otimes Y_{jC}^D \right) \\
 &= \bar{q} \cdot \left( \sum_{i=1}^n \sum_{j=1}^m c_i d_1 X_{iA}^B \otimes Y_{jC}^D \right) \\
 (5.54) \qquad &= c_1 d_1 = \|M_{1A}^B\| \|M_{2C}^D\|,
 \end{aligned}$$

which completes the proof that the norm (5.15) is additive. We can use this in the following manner to work out how our theory interacts with the tensor product method for combining systems. Firstly, consider the success probability over a tensor product

$$(5.55) \qquad S(M_{1A}^B \otimes M_{2C}^D) = \min_{\bar{q} \in \mathcal{Q}} \frac{\bar{q}_{BD}^{AC} \bullet M_{1A}^B \otimes M_{2C}^D}{\|M_{1A}^B \otimes M_{2C}^D\|}$$

which we know is equal to

$$(5.56) \qquad \min_{\bar{q} \in \mathcal{Q}} \frac{\bar{q}_{BD}^{AC} \bullet M_{1A}^B \otimes M_{2C}^D}{\|M_{1A}^B\| \|M_{2C}^D\|}.$$

Note that

$$\begin{aligned}
 \min_{\bar{q} \in \mathcal{Q}} \bar{q}_{BD}^{AC} \bullet M_{1A}^B \otimes M_{2C}^D &= \min_{\rho^{AC}} \rho^{AC} \bullet M_{1A}^B \otimes M_{2C}^D \bullet I_B \bullet I_D \\
 &= \min_{\rho^{AC}} \rho^{AC} \bullet E_{1A} \otimes E_{2C} \\
 (5.57) \qquad &= \min_{\rho} \text{Tr}[\rho(E_1 \otimes E_2)].
 \end{aligned}$$

Since  $E_1 \otimes E_2$  is product, we can perform the minimisation as a product also,

$$(5.58) \qquad = \min_{\rho_1} \text{Tr}[\rho_1 E_1] \min_{\rho_2} \text{Tr}[\rho_2 E_2].$$

Hence

$$(5.59) \qquad S(M_{1A}^B \otimes M_{2C}^D) = S(M_{1A}^B) S(M_{2C}^D).$$

Therefore, as the success probability is one minus the resource we have the relation

$$(5.60) \qquad 1 - R(M_{1A}^B \otimes M_{2C}^D) = (1 - R(M_{1A}^B))(1 - R(M_{2C}^D)).$$

for all channels that have the form described in result 5.5.

## 5.5 Discussion and Conclusion

In this chapter, we have discussed the resource theory of post-selection. In light of quantum resource theories (which the resource theory of post-selection goes beyond) our results indicate

some fundamental similarities and differences between this new type of resource and well studied quantum resources. For instance, in our theory, only a weaker version of convexity and additivity hold than in many quantum resource theories.

We have given a few illustrative examples for how to calculate quantities in this resource theory, and proofs of all of the properties claimed. Further questions remain however, for instance

- How well does the resource theory of post-selection play with other resource theories. In the resource theory of post-selection, the free set is all of quantum theory. How does the theory change when one introduces further restrictions on the free set and operations? And how can we trade the resource of post-selection for other resources?
- Can we use the resource theory of post-selection to characterise quantum, and post-quantum processes? For instance, a natural question is given a quantum process that violates a causal inequality, how much post-selection does one need to implement this process in an experiment using quantum theory? Are there any fundamental limitations on certain types of causal inequalities that require more post-selection to implement than other types of causal inequalities?
- Can we use the resource theory of post-selection to better characterise experiments across the field of quantum information?
- How are the weak measurement and weak value related to the resource theory of post-selection? Can we obtain a natural relation between this resource and information theoretic quantities related to measurement, say, the information-disturbance trade off?
- Does the resource theory of post-selection provide a natural way to compare quantum algorithms that use post-selection in different ways, in a complexity theoretic sense?

The resource theory of post-selection also has broadly raised some interesting questions concerning post-selection. Consider an unsupervised machine learning programme. In this programme, the machine attempts to solve a task by trial and error, and after a generation of trial and error experiments, the best solutions are used to birth a new generation of problem solvers. After many generations, the programme has learned how to solve the problem. Clearly, post-selection has not been used in this task, but feed-forward control on the programming of subsequent generations from prior ones. What, if any, advantage can post-selection provide to supervised machine learning tasks?

The pseudo-density matrix formalism for quantum mechanics [249] is a formalism where correlations can imply that events are time-like, not space-like related. Bell inequalities and causal inequalities can be violated with a pseudo-density matrix. A shortcoming of this formalism is that the implication is one way, a correlation cannot tell you if an event is space-like or time-like uniquely. Can we characterise a similar resource theory of ‘density matrix pseudo-nality’?

Also raised is the questions of where in an experiment does one need to post-select to obtain maximum ‘advantage’. The  $\bullet$  operation in some ways location neutral, indices are just being contracted. Also, all the quantum objects in a given process are future identity preserving. Does this mean we can harness the full power of post-selection by moving all of the measurements that are being post-selected to earlier stages of the process? If not, why not? Further to this point, we would also be interested to investigate the local power of post-selection. Supposed the contraction in the definition (5.10) are not forced to be total, and there could be a post-selection being performed by a single party only. We have went some way into investigating this by analysing the tensor product structure of the resource theory of post-selection, however an operational task which has a different structure under a local versus a global post-selection would be interesting to investigate - especially if one considers the foundational question of do black holes post-select on quantum states.

Another point is that all examples in this chapter are fairly trivial. Are there less trivial examples of post-selections which lead to analytic solutions to equation (5.10). A tangential but related question is ‘what is the exact function of quantum mechanics in the resource theory of post-selection?’. Where exactly is entanglement playing an important role?

We speculate that post-selection is a science of purposely being ignored. As a result, anyone seeking to increase the efficiency of some practice based upon a numerical model that pertains to real experiments with real inefficiencies ought to be aware of its deadly perils. However, the trickster may yet find its usage in fields as wide as land-use, the power industry, census taking, financial modelling, data science at large, fabrication, machining, or anywhere where there is a *biased* measurement inefficiency present.

Finally, is there any resource that can be spent to simulate post-selection. Post-selection is enormously powerful, but what, if anything, can simulate it in full generality.

The physical principle in this chapter is erasure. Information is physical, and erasing it comes at a cost.



## CONCLUSIONS

**B**egins a chapter, and ends a Thesis. In this section, we will review our main conclusions from each chapter, speculate on some directions for future research, and ask some open questions.

## 6.1 Discussions from Thermodynamics

In Chapter 2, we discussed the implicit and explicit battery formalism for quantum thermodynamics. We outlined results from prior work for the case of unitary work extraction, and then generalised this to other types of quantum evolution - quantum channels and quantum measurements. We showed that catalytic channels are not useful for work extraction, and that unital channels are also not useful for work extraction, that being useless for work extraction means that the channel is unital but that unital channels do not necessarily have a catalytic implementation. We also introduced a thermodynamic work for the case-by-case basis where measurement outcomes are obtained, and called this the conditional work extraction. This quantity depended upon the state of the machine that was used to implement it (and so could not hope to be derived in the implicit formalism) and also provided insight for the final section, making sense of post-selected quantum thermodynamics. Post-selected quantum thermodynamics is very interesting, one can use a genuinely post-selected quantum channel to provide an unbounded amount of work, and we give a protocol for doing so.

### 6.1.1 Further Works

Many open questions remain. I would love to see more realisations of quantum thermodynamic experiments. There have been very exciting proposals, for instance, using a molecule of DNA



as a thermal engine. How far does this paradigm go? Do quantum thermodynamic experiments permit the construction of nanoscale thermal machines that do useful things? An application of refrigeration to alter the temperature of an enzyme active site would allow an unprecedented level of control over the biological nano-regime. Could we use quantum heat engines to cool classical computers at an unprecedented efficiency?

Also, as I mentioned, the use of entanglement to assist metrology is an avenue that is begging for an exploration, although first studies are appearing [250]. This method would have no classical analogue, making it attractive to both popular science and academics alike. Perhaps, it would also even be useful in industry, but I cannot profess to know such things.

## 6.2 Discussions from Causality

In chapter 3 we prove that coherent quantum control over measurement order does not permit one to violate a causal inequality. We do this, not in the process matrix formalism or the two-time formalism, but only using ‘vanilla’ quantum theory. This result is interesting, it implies for instance that we will not see causal inequality violation in the outcome statistics of a quantum computation for example. Now, I don’t believe that anybody prior to this work thought that we would - however proving it is important. Moreover, we now have a protocol which will allow us to transform a process which is realised in quantum theory by a superposition of causal orders into one which has a classical causal model.

In chapter 4 we show that a local notion of causality - the no-backwards-in-time-signalling conditions are insufficient to ensure a globally causal explanation for any correlation. This is a distinctly tripartite effect, it has been shown that in the bipartite (and trivially, the single party) case, then the no-backwards-in-time-signalling conditions do indeed ensure the correlations have a causal explanation. Classifying these effects is important, it is well known that in the field of non-locality there are distinctly tripartite effects (such as the achievable correlations when sharing a GHZ state).

### 6.2.1 Further Works

I would love to investigate further into the causal framework. A natural question one can ask is, much like we did in Chapter 4, how more and more players entering the game changes things, and the types of causal structures that emerge. For instance, it has been well noted that non-locality in network type scenarios can exhibit a vastly different character to the standard two-input, two-output, two player scenario of Bell [251]. Of course, the classically causal literature of pearl and others [58] is very concerned with networks. How quantum theory plays with these constraints exactly is an interesting problem.

We have hardly mentioned relativistic causality in this work. We instead favour calculating in non-relativistic but explicit models. Do our results change when relativity is in the game?

Moreover, what is the physical principle that enforces quantum theory to not violate causal inequalities. We know that it does not, but we would like an explanation for this fact.

## 6.3 Discussions from Post-Selection

In chapter 5 we have developed a resource theory to study post-selection. We quantify the ‘minimum amount’ of post-selection needed to implement a given post-selected measurement and shown that this object is a resource monotone.

We have given some examples which outline how to calculate in this theory, and some more examples for how the theory interacts with our usual method for combining systems, the tensor product. We hope that these are illustrative.

### 6.3.1 Further works

Quantify, quantify, quantify! As mentioned, lots of algorithms, computation and experiments utilise post-selection all the time. How much they are utilising, and is it being done in the most efficient fashion is the subject of another thesis that I will not be the one to write.

How does post-selection play with locality? In a multiparty setting, are there post-selections that acting only locally can be made to do something useful? Are there types of tasks where one needs a global post selection to glean any advantage in?

Finally, the post-selected world is full of paradoxes and fantastical effects. Can we use this work to better understand quantum Cheshire cats, weak measurements or the physics two-time states?

## 6.4 Outro

This thesis clearly does not represent the sum total of the field of quantum information science, and does not claim too. What is the end goal for quantum information science is an imprecise question. As the field evolves, and in the light of quantum technologies making an ever more transformative impact upon the world, we are free to define our goal as we wish. At the current moment for instance, there is much excitement surrounding using quantum mechanical systems to process information and do quantum computations. Steps to making a fully fault tolerant quantum computer are taken in research across many disciplines of science and engineering, however what benefits quantum computing can have to the world, will be described by quantum information science.

Today, we have a better understanding of the universe than ever before. We understand how the atoms of physics build molecules of chemistry which build the organisms of biology which build life. We understand how these atoms were formed during the supernova of distant stars.

The quantum realm however, and all that goes with it, still carries an air of mystery. Hopefully this work will represent an incremental step towards better understanding the quantum.

## BIBLIOGRAPHY

- [1] Yakir Aharonov and Daniel Rohrlich.  
“Quantum Paradoxes: Quantum Theory for the Perplexed”.  
In: *Wiley-VCH* (Sept. 2003).  
DOI: 10.1002/9783527619115.
- [2] A. Einstein.  
“Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt”.  
In: *Annalen der Physik* 322.6 (1905).  
DOI: <https://doi.org/10.1002/andp.19053220607>.
- [3] J. Larmor.  
“LXIII. On the theory of the magnetic influence on spectra; and on the radiation from moving ions”.  
In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 44.271 (1897).  
DOI: 10.1080/14786449708621095.
- [4] N. Bohr.  
“I. On the constitution of atoms and molecules”.  
In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 26.151 (1913).  
DOI: 10.1080/14786441308634955.
- [5] Louis de Broglie.  
“XXXV. A tentative theory of light quanta”.  
In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 47.278 (1924).  
DOI: 10.1080/14786442408634378.
- [6] E. Noether.  
“Invariante Variationsprobleme”.  
In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918).  
URL: <http://eudml.org/doc/59024>.

- [7] Lucien Hardy.  
*Quantum Theory From Five Reasonable Axioms*.  
2001.  
arXiv: quant-ph/0101012 [quant-ph].
- [8] Gabriele Carcassi, Lorenzo Maccone, and Christine A. Aidala.  
“Four Postulates of Quantum Mechanics Are Three”.  
In: *Physical Review Letters* 126.11 (Mar. 2021).  
DOI: 10.1103/physrevlett.126.110402.
- [9] Michael A. Nielsen and Isaac L. Chuang.  
*Quantum Computation and Quantum Information: 10th Anniversary Edition*.  
Cambridge University Press, 2011.
- [10] Walter Rudin.  
*Real and Complex Analysis, 3rd Ed.*  
McGraw-Hill, Inc., 1987.
- [11] J. Preskill.  
*Lecture Notes for Physics 229: Quantum Information and Computation*.  
CreateSpace Independent Publishing Platform, 2015.  
URL: <https://books.google.co.uk/books?id=MIv8rQEACAAJ>.
- [12] John Goold et al.  
“The role of quantum information in thermodynamics—a topical review”.  
In: *Journal of Physics A: Mathematical and Theoretical* 49.14 (2016).  
DOI: 10.1088/1751-8113/49/14/143001.
- [13] Patrick P. Potts.  
*Introduction to Quantum Thermodynamics (Lecture Notes)*.  
2019.  
arXiv: 1906.07439 [quant-ph].
- [14] Sebastian Deffner and Steve Campbell.  
*Quantum Thermodynamics: An introduction to the thermodynamics of quantum information*.  
2019.  
arXiv: 1907.01596 [quant-ph].
- [15] Philipp Strasberg and Andreas Winter.  
*First and Second Law of Quantum Thermodynamics: A Consistent Derivation Based on a Microscopic Definition of Entropy*.  
2021.  
arXiv: 2002.08817 [quant-ph].

- 
- [16] Yelena Guryanova, Nicolai Friis, and Marcus Huber.  
“Ideal Projective Measurements Have Infinite Resource Costs”.  
In: *Quantum* 4 (Jan. 2020).  
DOI: 10.22331/q-2020-01-13-222.
- [17] A.S. Eddington.  
*New Pathways in Science*.  
Read Books, 2007.  
ISBN: 9781406740868.
- [18] David Deutsch and Chiara Marletto.  
“Constructor theory of information”.  
In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 471.2174 (Feb. 2015).  
DOI: 10.1098/rspa.2014.0540.
- [19] Paul Skrzypczyk, Anthony J. Short, and Sandu Popescu.  
“Work extraction and thermodynamics for individual quantum systems”.  
In: *Nature Communications* 5.1 (2014).  
DOI: 10.1038/ncomms5185.
- [20] Benjamin Schumacher.  
“Quantum coding”.  
In: *Phys. Rev. A* 51 (4 Apr. 1995).  
DOI: 10.1103/PhysRevA.51.2738.
- [21] Fumio Hiai and Dénes Petz.  
“The proper formula for relative entropy and its asymptotics in quantum probability”.  
In: *Communications in Mathematical Physics* 143.1 (Dec. 1991).  
DOI: 10.1007/BF02100287.
- [22] Charles H. Bennett et al.  
“Concentrating partial entanglement by local operations”.  
In: *Phys. Rev. A* 53 (4 1996).  
DOI: 10.1103/PhysRevA.53.2046.
- [23] Fernando G. S. L. Brandão et al.  
“Resource Theory of Quantum States Out of Thermal Equilibrium”.  
In: *Phys. Rev. Lett.* 111 (25 Dec. 2013).  
DOI: 10.1103/PhysRevLett.111.250404.
- [24] Nilanjana Datta.  
“Min- and Max-Relative Entropies and a New Entanglement Monotone”.  
In: *IEEE Transactions on Information Theory* 55.6 (June 2009).  
DOI: 10.1109/tit.2009.2018325.

- [25] Mario Berta, Kaushik P. Seshadreesan, and Mark M. Wilde.  
“Rényi generalizations of the conditional quantum mutual information”.  
In: *Journal of Mathematical Physics* 56.2 (2015).  
DOI: 10.1063/1.4908102.
- [26] Mark M. Wilde, Andreas Winter, and Dong Yang.  
“Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Rényi Relative Entropy”.  
In: *Communications in Mathematical Physics* 331.2 (Oct. 2014).  
DOI: 10.1007/s00220-014-2122-x.
- [27] A. Einstein, B. Podolsky, and N. Rosen.  
“Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?”  
In: *Phys. Rev.* 47 (10 May 1935).  
DOI: 10.1103/PhysRev.47.777.
- [28] N. Bohr.  
“Can Quantum-Mechanical Description of Physical Reality be Considered Complete?”  
In: *Phys. Rev.* 48 (8 Oct. 1935).  
DOI: 10.1103/PhysRev.48.696.
- [29] J. S. Bell.  
“On the Einstein Podolsky Rosen paradox”.  
In: *Physics Physique Fizika* 1 (3 1964).  
DOI: 10.1103/PhysicsPhysiqueFizika.1.195.
- [30] Jian-Wei Pan et al.  
“Multiphoton entanglement and interferometry”.  
In: *Rev. Mod. Phys.* 84 (2 May 2012).  
DOI: 10.1103/RevModPhys.84.777.
- [31] Jan-Åke Larsson.  
“Loopholes in Bell inequality tests of local realism”.  
In: *Journal of Physics A: Mathematical and Theoretical* 47.42 (Oct. 2014).  
DOI: 10.1088/1751-8113/47/42/424003.
- [32] David I. Kaiser.  
*Tackling Loopholes in Experimental Tests of Bell’s Inequality*.  
2020.  
arXiv: 2011.09296 [quant-ph].
- [33] Patrick Thomas Fraser and Barry C Sanders.  
“Loophole-Free Bell Tests and the Falsification of Local Realism”.  
In: *Journal of Student Science and Technology* 10 (Aug. 2017).  
DOI: 10.13034/jsst.v10i1.164.

- 
- [34] John F. Clauser et al.  
“Proposed Experiment to Test Local Hidden-Variable Theories”.  
In: *Phys. Rev. Lett.* 23 (15 Oct. 1969).  
DOI: 10.1103/PhysRevLett.23.880.
- [35] Nicolas Brunner et al.  
“Bell nonlocality”.  
In: *Rev. Mod. Phys.* 86 (2 Apr. 2014).  
DOI: 10.1103/RevModPhys.86.419.
- [36] J F Clauser and A Shimony.  
“Bell’s theorem. Experimental tests and implications”.  
In: *Reports on Progress in Physics* 41.12 (Dec. 1978).  
DOI: 10.1088/0034-4885/41/12/002.
- [37] B. S. Cirel’son.  
“Quantum generalizations of Bell’s inequality”.  
In: *Letters in Mathematical Physics* 4.2 (Mar. 1980).  
DOI: 10.1007/BF00417500.
- [38] Sandu Popescu and Daniel Rohrlich.  
“Quantum nonlocality as an axiom”.  
In: *Foundations of Physics* 24.3 (1994).  
DOI: 10.1007/BF02058098.
- [39] Marcin Pawłowski et al.  
“Information causality as a physical principle”.  
In: *Nature* 461 (Oct. 2009).  
DOI: 10.1038/nature08400.
- [40] W. Pauli.  
*General Principles of Quantum Mechanics*.  
Springer, Berlin Heidelberg, 1980.
- [41] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone.  
“Quantum time”.  
In: *Physical Review D* 92.4 (Aug. 2015).  
DOI: 10.1103/physrevd.92.045033.
- [42] Don N. Page and William K. Wootters.  
“Evolution without evolution: Dynamics described by stationary observables”.  
In: *Phys. Rev. D* 27 (12 June 1983).  
DOI: 10.1103/PhysRevD.27.2885.
- [43] Bryce S. DeWitt.



- “Quantum Theory of Gravity. I. The Canonical Theory”.  
In: *Phys. Rev.* 160 (5 Aug. 1967).  
DOI: 10.1103/PhysRev.160.1113.
- [44] Carlo Rovelli.  
“Loop Quantum Gravity”.  
In: *Living Reviews in Relativity* 1.1 (Jan. 1998).  
DOI: 10.12942/lrr-1998-1.
- [45] C. Marletto and V. Vedral.  
“Evolution without evolution and without ambiguities”.  
In: *Phys. Rev. D* 95 (Feb. 2017).  
DOI: 10.1103/PhysRevD.95.043510.
- [46] Norbert Grot, Carlo Rovelli, and Ranjeet S. Tate.  
“Time of arrival in quantum mechanics”.  
In: *Physical Review A* 54.6 (Dec. 1996).  
DOI: 10.1103/physreva.54.4676.
- [47] Paul Erker et al.  
“Autonomous Quantum Clocks: Does Thermodynamics Limit Our Ability to Measure Time?”  
In: *Phys. Rev. X* 7 (3 Aug. 2017).  
DOI: 10.1103/PhysRevX.7.031022.
- [48] Mischa P. Woods, Ralph Silva, and Jonathan Oppenheim.  
“Autonomous Quantum Machines and Finite-Sized Clocks”.  
In: *Annales Henri Poincaré* 20.1 (Oct. 2018).  
DOI: 10.1007/s00023-018-0736-9.
- [49] Mischa P. Woods and Michał Horodecki.  
*The Resource Theoretic Paradigm of Quantum Thermodynamics with Control*.  
2019.  
arXiv: 1912.05562 [quant-ph].
- [50] Andreas Bauch.  
“Caesium atomic clocks: function, performance and applications”.  
In: *Measurement Science and Technology* 14.8 (July 2003).  
DOI: 10.1088/0957-0233/14/8/301.
- [51] Huzihiro Araki and Mutsuo M. Yanase.  
“Measurement of Quantum Mechanical Operators”.  
In: *Phys. Rev.* 120 (2 Oct. 1960).  
DOI: 10.1103/PhysRev.120.622.

- 
- [52] J. Oppenheim.  
“Quantum time”.  
In: *AIP Conf. Proc.* 493.1 (1999).  
DOI: 10.1063/1.1301593.
- [53] Yakir Aharonov, Peter G. Bergmann, and Joel L. Lebowitz.  
“Time Symmetry in the Quantum Process of Measurement”.  
In: *Phys. Rev.* 134 (6B June 1964).  
DOI: 10.1103/PhysRev.134.B1410.
- [54] Yakir Aharonov and Benni Reznik.  
““Weighing” a Closed System and the Time-Energy Uncertainty Principle”.  
In: *Physical Review Letters* 84 (July 1999).  
DOI: 10.1103/PhysRevLett.84.1368.
- [55] Yakir Aharonov and Lev Vaidman.  
*The Two-State Vector Formalism of Quantum Mechanics.*  
Berlin, Heidelberg, 2002.  
DOI: 10.1007/3-540-45846-8\_13.
- [56] Graciela De Pierris and Michael Friedman.  
“Kant and Hume on Causality”.  
In: *The Stanford Encyclopedia of Philosophy.*  
Winter 2018.  
Metaphysics Research Lab, Stanford University, 2018.
- [57] Daniel Kahneman.  
*Thinking, fast and slow.*  
Macmillan, 2011.
- [58] Judea Pearl.  
*Causality: Models, Reasoning and Inference.*  
USA: Cambridge University Press, 2009.  
ISBN: 052189560X.
- [59] Hans Reichenbach.  
*The Direction of Time.*  
Dover Publications, 1956.
- [60] John-Mark A. Allen et al.  
“Quantum Common Causes and Quantum Causal Models”.  
In: *Physical Review X* 7.3 (July 2017).  
DOI: 10.1103/physrevx.7.031021.
- [61] Jonathan Barrett, Robin Lorenz, and Ognjan Oreshkov.

- “Cyclic quantum causal models”.  
In: *Nature Communications* 12.1 (Feb. 2021).  
DOI: 10.1038/s41467-020-20456-x.
- [62] Cyril Branciard et al.  
“The simplest causal inequalities and their violation”.  
In: *New Journal of Physics* 18.1 (2015).  
DOI: 10.1088/1367-2630/18/1/013008.
- [63] Mafalda L. Almeida et al.  
“Guess Your Neighbor’s Input: A Multipartite Nonlocal Game with No Quantum Advantage”.  
In: *Physical Review Letters* 104.23 (June 2010).  
DOI: 10.1103/physrevlett.104.230404.
- [64] Tom Purves and Anthony J. Short.  
“Nonclassically causal correlations without backwards-in-time signaling”.  
In: *Physical Review A* 99.2 (Feb. 2019).  
DOI: 10.1103/physreva.99.022101.
- [65] Ä Baumeler and S. Wolf.  
“Perfect signaling among three parties violating predefined causal order”.  
In: *2014 IEEE International Symposium on Information Theory*.  
June 2014.  
DOI: 10.1109/ISIT.2014.6874888.
- [66] Yakir Aharonov et al.  
“Multiple-time states and multiple-time measurements in quantum mechanics”.  
In: *Phys. Rev. A* 79 (5 2009).  
DOI: 10.1103/PhysRevA.79.052110.
- [67] Yakir Aharonov, David Z. Albert, and Lev Vaidman.  
“How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100”.  
In: *Phys. Rev. Lett.* 60 (14 Apr. 1988).  
DOI: 10.1103/PhysRevLett.60.1351.
- [68] Charles H. Bennett et al.  
“Quantum nonlocality without entanglement”.  
In: *Phys. Rev. A* 59 (2 Feb. 1999).  
DOI: 10.1103/PhysRevA.59.1070.
- [69] Ognjan Oreshkov and Todd A. Brun.  
“Weak Measurements Are Universal”.  
In: *Phys. Rev. Lett.* 95 (11 Sept. 2005).

- DOI: 10.1103/PhysRevLett.95.110409.
- [70] Graeme Mitchison, Richard Jozsa, and Sandu Popescu.  
“Sequential weak measurement”.  
In: *Physical Review A* 76.6 (Dec. 2007).  
DOI: 10.1103/physreva.76.062105.
- [71] Alastair A. Abbott et al.  
“Anomalous Weak Values Without Post-Selection”.  
In: *Quantum* 3 (Oct. 2019).  
DOI: 10.22331/q-2019-10-14-194.
- [72] Charles H. Bennett.  
“Notes on Landauer’s principle, reversible computation, and Maxwell’s Demon”.  
In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 34.3 (2003).  
DOI: [https://doi.org/10.1016/S1355-2198\(03\)00039-X](https://doi.org/10.1016/S1355-2198(03)00039-X).
- [73] Bob Coecke and Aleks Kissinger.  
*Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning*.  
Cambridge University Press, 2017.  
DOI: 10.1017/9781316219317.
- [74] Jacob Biamonte.  
*Lectures on Quantum Tensor Networks*.  
2020.  
arXiv: 1912.10049 [quant-ph].
- [75] James Siciliano.  
*Rattlestar Ricklactica, Rick and Morty*.  
Dec. 2019.
- [76] Scott Aaronson.  
“Quantum computing, postselection, and probabilistic polynomial-time”.  
In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 461.2063 (2005).
- [77] Yenjo Han, Lane A. Hemaspaandra, and Thomas Thierauf.  
“Threshold computation and cryptographic security”.  
In: *Algorithms and Computation*.  
Berlin, Heidelberg: Springer Berlin Heidelberg, 1993.
- [78] Aram W. Harrow and Ashley Montanaro.  
“Quantum computational supremacy”.

## BIBLIOGRAPHY

---

- In: *Nature* 549 (Sept. 2017).  
DOI: 10.1038/nature23458.
- [79] Yenjo Han, Lane A. Hemaspaandra, and Thomas Thierauf.  
“Threshold Computation and Cryptographic Security”.  
In: *SIAM JOURNAL ON COMPUTING* 26 (1995).
- [80] Richard Beigel.  
“Perceptrons, PP, and the polynomial hierarchy”.  
In: *computational complexity* 4 (Dec. 1994).  
DOI: 10.1007/BF01263422.
- [81] Tom Purves and Anthony J. Short.  
“Channels, measurements, and postselection in quantum thermodynamics”.  
In: *Phys. Rev. E* 104 (1 July 2021).  
DOI: 10.1103/PhysRevE.104.014111.
- [82] Tom Purves and Anthony J. Short.  
“Quantum Theory Cannot Violate a Causal Inequality”.  
In: *Phys. Rev. Lett.* 127 (11 Sept. 2021).  
DOI: 10.1103/PhysRevLett.127.110402.
- [83] Patryk Lipka-Bartosik et al.  
“Operational Significance of the Quantum Resource Theory of Buscemi Nonlocality”.  
In: *PRX Quantum* 2 (Apr. 2021).  
DOI: 10.1103/prxquantum.2.020301.
- [84] J. Gemmer, M. Michel, and G. Mahler.  
*Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems*.  
2004.  
DOI: 10.1007/b98082.
- [85] Achim Schmidt.  
*Technical Thermodynamics for Engineers: Basics and Applications*.  
Jan. 2019.  
DOI: 10.1007/978-3-030-20397-9.
- [86] Dilip Kondepudi and Ilya Prigogine.  
“Basic Concepts and the Laws of Gases”.  
In: *Modern Thermodynamics* (2014).  
DOI: <https://doi.org/10.1002/9781118698723.ch1>.
- [87] Ronnie Kosloff.  
“Quantum Thermodynamics: A Dynamical Viewpoint”.

- In: *Entropy* 15.6 (2013).  
DOI: 10.3390/e15062100.
- [88] R Alicki.  
“The quantum open system as a model of the heat engine”.  
In: *Journal of Physics A: Mathematical and General* 12.5 (1979).  
DOI: 10.1088/0305-4470/12/5/007.
- [89] A. E Allahverdyan, R Balian, and Th. M Nieuwenhuizen.  
“Maximal work extraction from finite quantum systems”.  
In: *Europhysics Letters (EPL)* 67.4 (2004).  
DOI: 10.1209/epl/i2004-10101-2.
- [90] Armen E. Allahverdyan, Ramandeep S. Johal, and Guenter Mahler.  
“Work extremum principle: Structure and function of quantum heat engines”.  
In: *Phys. Rev. E* 77 (4 2008).  
DOI: 10.1103/PhysRevE.77.041118.
- [91] Noah Linden, Sandu Popescu, and Paul Skrzypczyk.  
“How Small Can Thermal Machines Be? The Smallest Possible Refrigerator”.  
In: *Phys. Rev. Lett.* 105 (13 2010).  
DOI: 10.1103/PhysRevLett.105.130401.
- [92] A. E. Allahverdyan and K. V. Hovhannisyan.  
“Work extraction from microcanonical bath”.  
In: *EPL (Europhysics Letters)* 95.6 (2011).  
DOI: 10.1209/0295-5075/95/60004.
- [93] Michał Horodecki and Jonathan Oppenheim.  
“Fundamental limitations for quantum and nanoscale thermodynamics”.  
In: *Nature Communications* 4.1 (2013).  
DOI: 10.1038/ncomms3059.
- [94] Janet Anders and Vittorio Giovannetti.  
“Thermodynamics of discrete quantum processes”.  
In: *New Journal of Physics* 15.3 (Mar. 2013).  
DOI: 10.1088/1367-2630/15/3/033022.
- [95] Jonathan G. Richens and Lluís Masanes.  
“Work extraction from quantum systems with bounded fluctuations in work”.  
In: *Nature Communications* 7.1 (2016).  
DOI: 10.1038/ncomms13511.
- [96] H. E. D. Scovil and E. O. Schulz-DuBois.  
“Three-Level Masers as Heat Engines”.

## BIBLIOGRAPHY

---

- In: *Phys. Rev. Lett.* 2 (6 1959).  
DOI: 10.1103/PhysRevLett.2.262.
- [97] J. E. Geusic, E. O. Schulz-DuBios, and H. E. D. Scovil.  
“Quantum Equivalent of the Carnot Cycle”.  
In: *Phys. Rev.* 156 (2 1967).  
DOI: 10.1103/PhysRev.156.343.
- [98] Friedemann Tonner and Günter Mahler.  
“Autonomous quantum thermodynamic machines”.  
In: *Phys. Rev. E* 72 (6 2005).  
DOI: 10.1103/PhysRevE.72.066118.
- [99] Artur S L Malabarba, Anthony J Short, and Philipp Kammerlander.  
“Clock-driven quantum thermal engines”.  
In: *New Journal of Physics* 17.4 (2015).  
DOI: 10.1088/1367-2630/17/4/045027.
- [100] Ronnie Kosloff and Amikam Levy.  
“Quantum Heat Engines and Refrigerators: Continuous Devices”.  
In: *Annual Review of Physical Chemistry* 65.1 (2014).  
DOI: 10.1146/annurev-physchem-040513-103724.
- [101] Sourav Bhattacharjee and Amit Dutta.  
*Quantum thermal machines and batteries*.  
2020.  
arXiv: 2008.07889 [quant-ph].
- [102] Peter Reimann.  
“Foundation of Statistical Mechanics under Experimentally Realistic Conditions”.  
In: *Phys. Rev. Lett.* 101 (19 Nov. 2008).  
DOI: 10.1103/PhysRevLett.101.190403.
- [103] Noah Linden et al.  
“Quantum mechanical evolution towards thermal equilibrium”.  
In: *Phys. Rev. E* 79 (6 2009).  
DOI: 10.1103/PhysRevE.79.061103.
- [104] Noah Linden et al.  
“Quantum mechanical evolution towards thermal equilibrium”.  
In: *Phys. Rev. E* 79 (6 2009).  
DOI: 10.1103/PhysRevE.79.061103.
- [105] Christian Gogolin and Jens Eisert.

- “Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems”.
- In: *Reports on Progress in Physics* 79.5 (2016).  
DOI: 10.1088/0034-4885/79/5/056001.
- [106] Yelena Guryanova et al.  
“Thermodynamics of quantum systems with multiple conserved quantities”.
- In: *Nature Communications* 7 (2016).  
DOI: 10.1038/ncomms12049.
- [107] Johan Åberg.  
“Catalytic Coherence”.
- In: *Physical Review Letters* 113.15 (2014).  
DOI: 10.1103/physrevlett.113.150402.
- [108] Michal Horodecki and Jonathan Oppenheim.  
“(Quantumness in the context of) Resource Theories”.
- In: *International Journal of Modern Physics B* 27.01n03 (2013).  
DOI: 10.1142/S0217979213450197.
- [109] Seth Lloyd.  
“Quantum-mechanical Maxwell’s demon”.
- In: *Phys. Rev. A* 56 (5 1997).  
DOI: 10.1103/PhysRevA.56.3374.
- [110] M. Esposito and C. Van den Broeck.  
“Second law and Landauer principle far from equilibrium”.
- In: *EPL (Europhysics Letters)* 95.4 (2011).  
DOI: 10.1209/0295-5075/95/40004.
- [111] Joan A. Vaccaro and Stephen M. Barnett.  
“Information erasure without an energy cost”.
- In: *Proceedings of the Royal Society A* 467.2130 (2011).  
DOI: 10.1098/rspa.2010.0577.
- [112] D. Janzing et al.  
“Thermodynamic Cost of Reliability and Low Temperatures: Tightening Landauer’s Principle and the Second Law”.
- In: *International Journal of Theoretical Physics* 39.12 (2000).  
DOI: 10.1023/A:1026422630734.
- [113] Günter Mahler, Jochen Gemmer, and Mathias Michel.  
“Emergence of thermodynamic behavior within composite quantum systems”.
- In: *Physica E: Low-dimensional Systems and Nanostructures* 29.1-2 (2005).  
DOI: 10.1016/j.physe.2005.05.001.



## BIBLIOGRAPHY

---

- [114] Gatién Verley and David Lacoste.  
“Fluctuation theorems and inequalities generalizing the second law of thermodynamics out of equilibrium”.  
In: *Phys. Rev. E* 86 (5 Nov. 2012).  
DOI: 10.1103/PhysRevE.86.051127.
- [115] Nicolas Brunner et al.  
“Virtual qubits, virtual temperatures, and the foundations of thermodynamics”.  
In: *Physical Review E* 85.5 (2012).  
DOI: 10.1103/physreve.85.051117.
- [116] Fernando Brandão et al.  
“The second laws of quantum thermodynamics”.  
In: *Proceedings of the National Academy of Sciences* 112.11 (2015).  
DOI: 10.1073/pnas.1411728112.
- [117] Lídia del Rio et al.  
“The thermodynamic meaning of negative entropy”.  
In: *Nature* 474 (2011).  
DOI: 10.1038/nature10123.
- [118] Oscar C O Dahlsten et al.  
“Inadequacy of von Neumann entropy for characterizing extractable work”.  
In: *New Journal of Physics* 13.5 (2011).  
DOI: 10.1088/1367-2630/13/5/053015.
- [119] Dario Egloff et al.  
*A measure of majorisation emerging from single-shot statistical mechanics.*  
2012.  
arXiv: 1207.0434 [quant-ph].
- [120] Fred I. Dretske.  
*Knowledge and the Flow of Information.*  
MIT Press, 1981.
- [121] J.C. Kieffer.  
“History of source coding”.  
In: *IEEE Information Theory Society Newsletter* 43 (June 1993).
- [122] C. E. Shannon.  
“A mathematical theory of communication”.  
In: *The Bell System Technical Journal* 27.3 (1948).  
DOI: 10.1002/j.1538-7305.1948.tb01338.x.
- [123] Eric Chitambar and Gilad Gour.

- “Quantum resource theories”.  
In: *Reviews of Modern Physics* 91.2 (Apr. 2019).  
DOI: 10.1103/revmodphys.91.025001.
- [124] Facundo Sapienza, Federico Cerisola, and Augusto J. Roncaglia.  
“Correlations as a resource in quantum thermodynamics”.  
In: *Nature Communications* 10.1 (June 2019).  
DOI: 10.1038/s41467-019-10572-8.
- [125] Martí Perarnau-Llobet et al.  
“Extractable Work from Correlations”.  
In: *Physical Review X* 5.4 (Oct. 2015).  
DOI: 10.1103/physrevx.5.041011.
- [126] James Ladyman, Stuart Presnell, and Anthony J. Short.  
“The use of the information-theoretic entropy in thermodynamics”. English.  
In: *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 39.2 (May 2008).  
DOI: 10.1016/j.shpsb.2007.11.004.
- [127] Luis Masanes and Jonathan Oppenheim.  
“A general derivation and quantification of the third law of thermodynamics”.  
In: *Nat. Commun.* 8 (2017).  
DOI: 10.1038/ncomms14538.
- [128] Manabendra N. Bera et al.  
“Generalized laws of thermodynamics in the presence of correlations”.  
In: *Nature Communications* 8.1 (Dec. 2017).  
DOI: 10.1038/s41467-017-02370-x.
- [129] David Jennings and Terry Rudolph.  
“Entanglement and the thermodynamic arrow of time”.  
In: *Phys. Rev. E* 81 (6 June 2010).  
DOI: 10.1103/PhysRevE.81.061130.
- [130] David Reeb and Michael M Wolf.  
“An improved Landauer principle with finite-size corrections”.  
In: *New Journal of Physics* 16.10 (Oct. 2014).  
DOI: 10.1088/1367-2630/16/10/103011.
- [131] Francesco Buscemi.  
“All Entangled Quantum States Are Nonlocal”.  
In: *Physical Review Letters* 108.20 (May 2012).  
DOI: 10.1103/physrevlett.108.200401.

- [132] Álvaro M. Alhambra et al.  
“Work and reversibility in quantum thermodynamics”.  
In: *Physical Review A* 97.6 (June 2018).  
DOI: 10.1103/physreva.97.062114.
- [133] Mark M. Wilde.  
“Recoverability in quantum information theory”.  
In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 471.2182 (Oct. 2015).  
DOI: 10.1098/rspa.2015.0338.
- [134] Naoki Nitta et al.  
“Li-ion battery materials: present and future”.  
In: *Materials Today* 18.5 (2015).  
DOI: <https://doi.org/10.1016/j.mattod.2014.10.040>.
- [135] “Science for Environment Policy: Towards the battery of the future”.  
In: *Brief produced for the European Commission DG Environment by the Science Communication Unit, UWE, Bristol* ().  
URL: <http://ec.europa.eu/science-environment-policy>.
- [136] Patryk Lipka-Bartosik, Paweł Mazurek, and Michał Horodecki.  
*Second law of thermodynamics for batteries with vacuum state*.  
2019.  
arXiv: 1905.12072 [quant-ph].
- [137] Robert Alicki et al.  
“Thermodynamics of Quantum Information Systems — Hamiltonian Description”.  
In: *Open Systems & Information Dynamics* 11.3 (Sept. 2004).  
DOI: 10.1023/B:OPSY.0000047566.72717.71.
- [138] Felix Binder et al.  
“Quantum thermodynamics of general quantum processes”.  
In: *Physical Review E* 91.3 (Mar. 2015).  
DOI: 10.1103/physreve.91.032119.
- [139] Laszlo Gyongyosi and Sandor Imre.  
*Properties of the Quantum Channel*.  
2012.  
arXiv: 1208.1270 [quant-ph].
- [140] Dan C. Marinescu and Gabriela M. Marinescu.  
“CHAPTER 3 - Classical and Quantum Information Theory”.  
In: *Classical and Quantum Information*.  
Boston: Academic Press, 2012.

- DOI: <https://doi.org/10.1016/B978-0-12-383874-2.00003-5>.
- [141] Heinz-Peter Breuer and Francesco Petruccione.  
“The Theory of Open Quantum Systems”.  
In:  
2002.  
DOI: 10.1093/acprof:oso/9780199213900.001.0001.
- [142] Xin Wang, Mark M Wilde, and Yuan Su.  
“Quantifying the magic of quantum channels”.  
In: *New Journal of Physics* 21.10 (2019).  
DOI: 10.1088/1367-2630/ab451d.
- [143] W McCutcheon et al.  
“Experimental demonstration of a measurement-based realisation of a quantum channel”.  
In: *New Journal of Physics* 20.3 (Mar. 2018).  
DOI: 10.1088/1367-2630/aa9b5c.
- [144] Graeme Smith.  
*Quantum Channel Capacities*.  
2010.  
arXiv: 1007.2855 [cs.IT].
- [145] A S Holevo.  
“Quantum coding theorems”.  
In: *Russian Mathematical Surveys* 53.6 (Dec. 1998).  
DOI: 10.1070/rm1998v053n06abeh000091.
- [146] H. P. Yuen.  
“Coding theorems of quantum information theory”.  
In: *Proceedings. 1998 IEEE International Symposium on Information Theory*.  
1998.  
DOI: 10.1109/ISIT.1998.708670.
- [147] K.J. Resch et al.  
“Distributing entanglement and single photons through an intra-city, free-space quantum channel”.  
In: *Opt. Express* 13.1 (Jan. 2005).  
DOI: 10.1364/OPEX.13.000202.
- [148] P Villorresi et al.  
“Experimental verification of the feasibility of a quantum channel between space and Earth”.  
In: *New Journal of Physics* 10.3 (2008).  
DOI: 10.1088/1367-2630/10/3/033038.

## BIBLIOGRAPHY

---

- [149] Paul Boes et al.  
“Von Neumann Entropy from Unitarity”.  
In: *Phys. Rev. Lett.* 122 (21 May 2019).  
DOI: 10.1103/PhysRevLett.122.210402.
- [150] Henrik Wilming.  
*Entropy and reversible catalysis*.  
2020.  
arXiv: 2012.05573 [quant-ph].
- [151] Christian B. Mendl and Michael M. Wolf.  
“Unital Quantum Channels – Convex Structure and Revivals of Birkhoff’s Theorem”.  
In: *Communications in Mathematical Physics* 289.3 (May 2009).  
DOI: 10.1007/s00220-009-0824-2.
- [152] Alexander Müller-Hermes and Christopher Perry.  
“All unital qubit channels are 4-noisy operations”.  
In: *Letters in Mathematical Physics* 109.1 (June 2018).  
DOI: 10.1007/s11005-018-1104-x.
- [153] William Gordon Ritter.  
“Quantum channels and representation theory”.  
In: *Journal of Mathematical Physics* 46.8 (2005).  
DOI: 10.1063/1.1945768.
- [154] N H Y Ng et al.  
“Limits to catalysis in quantum thermodynamics”.  
In: *New Journal of Physics* 17.8 (2015).  
DOI: 10.1088/1367-2630/17/8/085004.
- [155] Wim van Dam and Patrick Hayden.  
“Universal entanglement transformations without communication”.  
In: *Phys. Rev. A* 67 (6 June 2003).  
DOI: 10.1103/PhysRevA.67.060302.
- [156] Patryk Lipka-Bartosik and Paul Skrzypczyk.  
“All States are Universal Catalysts in Quantum Thermodynamics”.  
In: *Phys. Rev. X* 11 (1 Mar. 2021).  
DOI: 10.1103/PhysRevX.11.011061.
- [157] R. F. Werner and A. S. Holevo.  
“Counterexample to an additivity conjecture for output purity of quantum channels”.  
In: *Journal of Mathematical Physics* 43.9 (2002).  
DOI: 10.1063/1.1498491.

- 
- [158] F. Verstraete and H. Verschelde.  
“On quantum channels”.  
In: *arXiv: Quantum Physics* (2002).
- [159] Shmuel Friedland and Raphael Loewy.  
*On the extreme points of quantum channels*.  
2014.  
arXiv: 1309.5898 [math-ph].
- [160] Jon Magne Leinaas, Jan Myrheim, and Eirik Ovrum.  
“Geometrical aspects of entanglement”.  
In: *Phys. Rev. A* 74 (1 July 2006).  
DOI: 10.1103/PhysRevA.74.012313.
- [161] David Beckman et al.  
“Causal and localizable quantum operations”.  
In: *Phys. Rev. A* 64 (5 Oct. 2001).  
DOI: 10.1103/PhysRevA.64.052309.
- [162] Andrea Solfanelli et al.  
“Maximal energy extraction via quantum measurement”.  
In: *Journal of Statistical Mechanics: Theory and Experiment* 2019.9 (Sept. 2019).  
DOI: 10.1088/1742-5468/ab3721.
- [163] Augusto J. Roncaglia, Federico Cerisola, and Juan Pablo Paz.  
“Work Measurement as a Generalized Quantum Measurement”.  
In: *Phys. Rev. Lett.* 113 (25 Dec. 2014).  
DOI: 10.1103/PhysRevLett.113.250601.
- [164] S. Suomela et al.  
“Moments of work in the two-point measurement protocol for a driven open quantum system”.  
In: *Physical Review B* 90.9 (Sept. 2014).  
DOI: 10.1103/physrevb.90.094304.
- [165] María García Díaz, Giacomo Guarneri, and Mauro Paternostro.  
*Quantum work statistics with initial coherence*.  
2020.  
arXiv: 2007.00042 [quant-ph].
- [166] C. Jarzynski.  
“Nonequilibrium Equality for Free Energy Differences”.  
In: *Phys. Rev. Lett.* 78 (14 Apr. 1997), pp. 2690–2693.  
DOI: 10.1103/PhysRevLett.78.2690.  
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.78.2690>.

## BIBLIOGRAPHY

---

- [167] Michele Campisi, Peter Hänggi, and Peter Talkner.  
“Colloquium: Quantum fluctuation relations: Foundations and applications”.  
In: *Rev. Mod. Phys.* 83 (3 July 2011).  
DOI: 10.1103/RevModPhys.83.771.
- [168] Yunong Shi, Christopher Chamberland, and Andrew Cross.  
“Fault-tolerant preparation of approximate GKP states”.  
In: *New Journal of Physics* 21.9 (Sept. 2019).  
DOI: 10.1088/1367-2630/ab3a62.
- [169] Yoon-Ho Kim, Sergei P. Kulik, and Yanhua Shih.  
“Quantum Teleportation of a Polarization State with a Complete Bell State Measurement”.  
In: *Phys. Rev. Lett.* 86 (7 Feb. 2001).  
DOI: 10.1103/PhysRevLett.86.1370.
- [170] Jürg Fröhlich and Baptiste Schubnel.  
“The preparation of states in quantum mechanics”.  
In: *Journal of Mathematical Physics* 57.4 (2016).  
DOI: 10.1063/1.4940696.
- [171] C.H. Bennett.  
“Demons, Engines and the Second Law”.  
In: *Scientific American* November (1987).
- [172] Yakir Aharonov, Sandu Popescu, and Daniel Rohrlich.  
“On conservation laws in quantum mechanics”.  
In: *Proceedings of the National Academy of Sciences* 118.1 (2021).  
DOI: 10.1073/pnas.1921529118.
- [173] Eliahu Cohen and Yakir Aharonov.  
“Quantum to Classical Transitions via Weak Measurements and Post-Selection”.  
In: *Quantum Structural Studies* (Nov. 2016).  
DOI: 10.1142/9781786341419\_0012.
- [174] Urmila Mahadev and Ronald de Wolf.  
*Rational approximations and quantum algorithms with postselection.*  
2014.  
arXiv: 1401.0912 [quant-ph].
- [175] Arun Kumar Pati.  
*Super Quantum Search Algorithm with Weak Value Amplification and Postselection.*  
2019.  
arXiv: 1910.12390 [quant-ph].
- [176] Julian Wechs et al.

- Quantum circuits with classical versus quantum control of causal order.*  
2021.  
arXiv: 2101.08796 [quant-ph].
- [177] Richard Feynman, Robert Leighton, and Matthew Sands.  
*The Feynman Lectures on Physics: Volume 3.*  
The Feynman Lectures on Physics.  
Boston: Addison-Wesley, 1963.
- [178] Giulio Chiribella et al.  
“Quantum computations without definite causal structure”.  
In: *Physical Review A* 88.2 (2013).  
DOI: 10.1103/physreva.88.022318.
- [179] Giulio Chiribella.  
“Perfect discrimination of no-signalling channels via quantum superposition of causal structures”.  
In: *Phys. Rev. A* 86 (4 Oct. 2012).  
DOI: 10.1103/PhysRevA.86.040301.
- [180] Cyril Branciard.  
“Witnesses of causal nonseparability: an introduction and a few case studies”.  
In: *Scientific Reports* 6.1 (May 2016).  
DOI: 10.1038/srep26018.
- [181] Magdalena Zych et al.  
“Bell’s theorem for temporal order”.  
In: *Nature Communications* 10.1 (Aug. 2019).  
DOI: 10.1038/s41467-019-11579-x.
- [182] Jonathan Barrett, Robin Lorenz, and Ognjan Oreshkov.  
*Quantum Causal Models.*  
2020.  
arXiv: 1906.10726 [quant-ph].
- [183] Lorenzo M. Procopio et al.  
“Experimental superposition of orders of quantum gates”.  
In: *Nature Communications* 6 (Aug. 2015).  
DOI: 10.1038/ncomms8913.
- [184] K. Goswami et al.  
“Indefinite Causal Order in a Quantum Switch”.  
In: *Phys. Rev. Lett.* 121 (9 2018).  
DOI: 10.1103/PhysRevLett.121.090503.



## BIBLIOGRAPHY

---

- [185] K. Goswami and J. Romero.  
“Experiments on quantum causality”.  
In: *AVS Quantum Science* 2.3 (Oct. 2020).  
DOI: 10.1116/5.0010747.
- [186] Aram W. Harrow et al.  
“Adaptive versus nonadaptive strategies for quantum channel discrimination”.  
In: *Phys. Rev. A* 81 (3 Mar. 2010).  
DOI: 10.1103/PhysRevA.81.032339.
- [187] Daniel Ebler, Sina Salek, and Giulio Chiribella.  
“Enhanced Communication with the Assistance of Indefinite Causal Order”.  
In: *Phys. Rev. Lett.* 120 (12 2018).  
DOI: 10.1103/PhysRevLett.120.120502.
- [188] Ognjan Oreshkov, Fabio Costa, and Časlav Brukner.  
“Quantum correlations with no causal order”.  
In: *Nature Communications* 3.1 (Oct. 2012).  
DOI: 10.1038/ncomms2076.
- [189] Časlav Brukner.  
“Quantum causality”.  
In: *Nature Physics* 10.4 (2014).  
DOI: 10.1038/nphys2930.
- [190] Alastair A. Abbott et al.  
“Multipartite causal correlations: Polytopes and inequalities”.  
In: *Physical Review A* 94.3 (Sept. 2016).  
DOI: 10.1103/physreva.94.032131.
- [191] Mateus Araújo et al.  
“Witnessing causal nonseparability”.  
In: *New Journal of Physics* 17.10 (Oct. 2015).  
DOI: 10.1088/1367-2630/17/10/102001.
- [192] Fabio Costa.  
*A no-go theorem for superpositions of causal orders.*  
2020.  
arXiv: 2008.06205 [quant-ph].
- [193] Simon Milz et al.  
*Quantum chicken-egg dilemmas: Delayed-choice causal order and the reality of causal non-separability.*  
2020.  
arXiv: 2008.07876 [quant-ph].

- [194] Y Aharonov and L Vaidman.  
“Complete description of a quantum system at a given time”.  
In: *Journal of Physics A: Mathematical and Theoretical* 24.10 (1991).  
DOI: 10.1088/0305-4470/24/10/018.
- [195] Yakir Aharonov et al.  
“Multiple-time states and multiple-time measurements in quantum mechanics”.  
In: *Phys. Rev. A* 79 (5 May 2009).  
DOI: 10.1103/PhysRevA.79.052110.
- [196] Ralph Silva et al.  
“Pre- and postselected quantum states: Density matrices, tomography, and Kraus operators”.  
In: *Phys. Rev. A* 89 (1 2014).  
DOI: 10.1103/PhysRevA.89.012121.
- [197] Peter Janotta and Haye Hinrichsen.  
“Generalized probability theories: what determines the structure of quantum theory?”  
In: *Journal of Physics A: Mathematical and Theoretical* 47.32 (July 2014).  
DOI: 10.1088/1751-8113/47/32/323001.
- [198] Yelena Guryanova et al.  
“Exploring the limits of no backwards in time signalling”.  
In: *Quantum* 3 (2019), p. 211.  
DOI: 10.22331/q-2019-12-09-211.
- [199] Ognjan Oreshkov and Christina Giarmatzi.  
“Causal and causally separable processes”.  
In: *New Journal of Physics* 18.9 (2016).  
DOI: 10.1088/1367-2630/18/9/093020.
- [200] Mateus Araújo et al.  
“A purification postulate for quantum mechanics with indefinite causal order”.  
In: *Quantum* 1 (2017).  
DOI: 10.22331/q-2017-04-26-10.
- [201] Giulio Chiribella, Giacomo Mauro D’Ariano, and Paolo Perinotti.  
“Probabilistic theories with purification”.  
In: *Phys. Rev. A* 81 (6 2010).  
DOI: 10.1103/PhysRevA.81.062348.
- [202] Adrien Feix, Mateus Araújo, and Časlav Brukner.  
“Causally nonseparable processes admitting a causal model”.  
In: *New Journal of Physics* 18.8 (2016).  
DOI: 10.1088/1367-2630/18/8/083040.

## BIBLIOGRAPHY

---

- [203] Pier A. Mello.  
“The von Neumann model of measurement in quantum mechanics”.  
In: *AIP Conference Proceedings* 1575.1 (2014).  
DOI: 10.1063/1.4861702.
- [204] Yakir Aharonov and Lev Vaidman.  
“Properties of a quantum system during the time interval between two measurements”.  
In: *Phys. Rev. A* 41 (1 Jan. 1990).  
DOI: 10.1103/PhysRevA.41.11.
- [205] L. Hardy.  
“Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure”.  
In: *Journal of Physics A Mathematical General* 40 (Mar. 2007).  
DOI: 10.1088/1751-8113/40/12/S12.  
eprint: gr-qc/0608043.
- [206] Lucien Hardy.  
“Quantum Gravity Computers: On the Theory of Computation with Indefinite Causal Structure”.  
In: *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle: Essays in Honour of Abner Shimony* (2009).  
DOI: 10.1007/978-1-4020-9107-0\_21.
- [207] Āmin Baumeler and Stefan Wolf.  
“Device-independent test of causal order and relations to fixed-points”.  
In: *New Journal of Physics* 18.3 (2016).  
DOI: 10.1088/1367-2630/18/3/035014.
- [208] Philippe Allard Guérin et al.  
“Exponential Communication Complexity Advantage from Quantum Superposition of the Direction of Communication”.  
In: *Physical Review Letters* 117.10 (Sept. 2016).  
DOI: 10.1103/physrevlett.117.100502.
- [209] Ralph Silva et al.  
“Connecting processes with indefinite causal order and multi-time quantum states”.  
In: *New Journal of Physics* 19.10 (2017).  
DOI: 10.1088/1367-2630/aa84fe.
- [210] D. Harlow.  
“Jerusalem lectures on black holes and quantum information”.  
In: *Rev. Mod. Phys.* 88 (1 2016).  
DOI: 10.1103/RevModPhys.88.015002.

- [211] P. C. W. Davies.  
“Quantum Weak Measurements and Cosmology”.  
In: *Quantum Theory: A Two-Time Success Story* (2014).  
DOI: 10.1007/978-88-470-5217-8\_7.
- [212] Susan Vineberg.  
“Dutch Book Arguments”.  
In: *The Stanford Encyclopedia of Philosophy*.  
Spring 2016.  
Metaphysics Research Lab, Stanford University, 2016.
- [213] Man-Duen Choi.  
“Completely positive linear maps on complex matrices”.  
In: *Linear Algebra and its Applications* 10.3 (1975).  
DOI: [https://doi.org/10.1016/0024-3795\(75\)90075-0](https://doi.org/10.1016/0024-3795(75)90075-0).
- [214] David Avis.  
“A Revised Implementation of the Reverse Search Vertex Enumeration Algorithm”.  
In:  
*Polytopes — Combinatorics and Computation*.  
Basel: Birkhäuser Basel, 2000.  
DOI: 10.1007/978-3-0348-8438-9\_9.
- [215] Ämin Baumeler, Amin Shiraz Gilani, and Jibran Rashid.  
*Unlimited non-causal correlations and their relation to non-locality*.  
2021.  
arXiv: 2104.06234 [quant-ph].
- [216] Seth Lloyd et al.  
“Closed Timelike Curves via Postselection: Theory and Experimental Test of Consistency”.  
In: *Phys. Rev. Lett.* 106 (4 2011).  
DOI: 10.1103/PhysRevLett.106.040403.
- [217] H. Snijders et al.  
“Purification of a single-photon nonlinearity”.  
In: *Nature Communications* 7.1 (Aug. 2016).  
DOI: 10.1038/ncomms12578.
- [218] Jaromír Fiurášek and Nicolas J. Cerf.  
“Gaussian postselection and virtual noiseless amplification in continuous-variable quantum key distribution”.  
In: *Physical Review A* 86.6 (Dec. 2012).  
DOI: 10.1103/physreva.86.060302.
- [219] Shengshi Pang and Todd A. Brun.

## BIBLIOGRAPHY

---

- “Improving the Precision of Weak Measurements by Postselection Measurement”.  
In: *Physical Review Letters* 115.12 (Sept. 2015).  
DOI: 10.1103/physrevlett.115.120401.
- [220] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone.  
“Advances in quantum metrology”.  
In: *Nature Photonics* 5.4 (Apr. 2011).  
DOI: 10.1038/nphoton.2011.35.
- [221] Saki Tanaka and Naoki Yamamoto.  
“Information amplification via postselection: A parameter-estimation perspective”.  
In: *Physical Review A* 88.4 (Oct. 2013).  
DOI: 10.1103/physreva.88.042116.
- [222] Eliahu Cohen.  
“What Weak Measurements and Weak Values Really Mean: Reply to Kastner”.  
In: *Foundations of Physics* 47.10 (Oct. 2017).  
DOI: 10.1007/s10701-017-0107-2.
- [223] F. Piacentini et al.  
“Measuring Incompatible Observables by Exploiting Sequential Weak Values”.  
In: *Phys. Rev. Lett.* 117 (17 Oct. 2016).  
DOI: 10.1103/PhysRevLett.117.170402.
- [224] F. Piacentini et al.  
“Experiment Investigating the Connection between Weak Values and Contextuality”.  
In: *Phys. Rev. Lett.* 116 (18 May 2016).  
DOI: 10.1103/PhysRevLett.116.180401.
- [225] Lev Vaidman et al.  
“Weak value beyond conditional expectation value of the pointer readings”.  
In: *Physical Review A* 96.3 (Sept. 2017).  
DOI: 10.1103/physreva.96.032114.
- [226] Sang Kyu Kwak and Jong Hae Kim.  
“Statistical data preparation: management of missing values and outliers”.  
In: *Korean journal of anesthesiology* 70.4 (Aug. 2017).  
DOI: 10.4097/kjae.2017.70.4.407.
- [227] Richard P. Feynman and Richard Phillips.  
*Surely you're joking, Mr. Feynman : adventures of a curious character.*  
First edition. New York : W.W. Norton, 1984, c1985., 1985.
- [228] Philippe Faist et al.  
“The minimal work cost of information processing”.

- In: *Nature Communications* 6.1 (2015).  
DOI: 10.1038/ncomms8669.
- [229] Michael H. Freedman, Matthew B. Hastings, and Modjtaba Shokrian Zini.  
*Symmetry Protected Quantum Computation*.  
2021.  
arXiv: 2105.04649 [quant-ph].
- [230] Ryszard Horodecki et al.  
“Quantum entanglement”.  
In: *Rev. Mod. Phys.* 81 (2 June 2009).  
DOI: 10.1103/RevModPhys.81.865.
- [231] Alexander Streltsov, Gerardo Adesso, and Martin B. Plenio.  
“Colloquium: Quantum coherence as a resource”.  
In: *Rev. Mod. Phys.* 89 (4 Oct. 2017).  
DOI: 10.1103/RevModPhys.89.041003.
- [232] Stephen D. Bartlett, Terry Rudolph, and Robert W. Spekkens.  
“Reference frames, superselection rules, and quantum information”.  
In: *Rev. Mod. Phys.* 79 (2 Apr. 2007).  
DOI: 10.1103/RevModPhys.79.555.
- [233] Christian Weedbrook et al.  
“Gaussian quantum information”.  
In: *Rev. Mod. Phys.* 84 (2 May 2012).  
DOI: 10.1103/RevModPhys.84.621.
- [234] Barbara Amaral.  
*Resource Theory of Contextuality*.  
2019.  
arXiv: 1904.04182 [quant-ph].
- [235] Fernando G. S. L. Brandão and Martin B. Plenio.  
“Entanglement theory and the second law of thermodynamics”.  
In: *Nature Physics* 4.11 (Nov. 2008).  
DOI: 10.1038/nphys1100.
- [236] Fernando G. S. L. Brandão and Martin B. Plenio.  
“A Reversible Theory of Entanglement and its Relation to the Second Law”.  
In: *Communications in Mathematical Physics* 295.3 (May 2010).  
DOI: 10.1007/s00220-010-1003-1.
- [237] M. B. Plenio.  
“Logarithmic Negativity: A Full Entanglement Monotone That is not Convex”.

## BIBLIOGRAPHY

---

- In: *Physical Review Letters* 95.9 (Aug. 2005).  
DOI: 10.1103/physrevlett.95.090503.
- [238] Matteo Lostaglio.  
“An introductory review of the resource theory approach to thermodynamics”.  
In: *Reports on Progress in Physics* 82.11 (2019).  
DOI: 10.1088/1361-6633/ab46e5.
- [239] Carlos Palazuelos.  
“Superactivation of Quantum Nonlocality”.  
In: *Phys. Rev. Lett.* 109 (19 Nov. 2012).  
DOI: 10.1103/PhysRevLett.109.190401.
- [240] T. S. Cubitt, J. Chen, and A. W. Harrow.  
“Superactivation of the Asymptotic Zero-Error Classical Capacity of a Quantum Channel”.  
In: *IEEE Transactions on Information Theory* 57.12 (2011).  
DOI: 10.1109/TIT.2011.2169109.
- [241] Barbara Synak-Radtke and Michał Horodecki.  
“On asymptotic continuity of functions of quantum states”.  
In: *Journal of Physics A: Mathematical and Theoretical* (2006).  
DOI: 10.1088/0305-4470/39/26/102.
- [242] M. Fannes.  
“A continuity property of the entropy density for spin lattice systems”.  
In: *Communications in Mathematical Physics* (Dec. 1973).  
DOI: 10.1007/BF01646490.
- [243] Koenraad M R Audenaert.  
“A sharp continuity estimate for the von Neumann entropy”.  
In: *Journal of Physics A: Mathematical and Theoretical* (2007).  
DOI: 10.1088/1751-8113/40/28/s18.
- [244] Armin Tavakoli et al.  
“Correlations in star networks: from Bell inequalities to network inequalities”.  
In: *New Journal of Physics* 19.7 (2017).  
DOI: 10.1088/1367-2630/aa7673.
- [245] Marc-Olivier Renou et al.  
“Genuine Quantum Nonlocality in the Triangle Network”.  
In: *Physical Review Letters* 123.14 (Sept. 2019).  
DOI: 10.1103/physrevlett.123.140401.
- [246] Eric Chitambar et al.  
“Everything You Always Wanted to Know About LOCC (But Were Afraid to Ask)”.

- 
- In: *Communications in Mathematical Physics* 328.1 (Mar. 2014).  
DOI: 10.1007/s00220-014-1953-9.
- [247] Sevag Gharibian.  
*Strong NP-Hardness of the Quantum Separability Problem*.  
2009.  
arXiv: 0810.4507 [quant-ph].
- [248] Yakir Aharonov et al.  
“Quantum Cheshire Cats”.  
In: *New Journal of Physics* 15.11 (Nov. 2013).  
DOI: 10.1088/1367-2630/15/11/113015.
- [249] Joseph Fitzsimons, Jonathan Jones, and Vlatko Vedral.  
*Quantum correlations which imply causation*.  
2013.  
arXiv: 1302.2731 [quant-ph].
- [250] Zixin Huang, Chiara Macchiavello, and Lorenzo Maccone.  
“Usefulness of entanglement-assisted quantum metrology”.  
In: *Physical Review A* 94.1 (July 2016).  
DOI: 10.1103/physreva.94.012101.
- [251] Nicolas Gisin et al.  
“Constraints on nonlocality in networks from no-signaling and independence”.  
In: *Nature Communications* 11.1 (May 2020).  
DOI: 10.1038/s41467-020-16137-4.



