

Phase and interference phenomena in generalised probabilistic theories

A thesis submitted for the degree of
DPhil in Atomic & Laser Physics

by

Andrew J. P. Garner

New College,
University of Oxford

Hilary Term, 2015



To my grandparents...

Phase and interference phenomena in generalised probabilistic theories

**Abstract of a thesis submitted for the degree of:
DPhil in Atomic & Laser Physics.**

Andrew J. P. Garner, New College, Hilary Term 2015

Phase lies at the heart of quantum physics and quantum information theory. A quantum bit is qualitatively different from a classical bit as it allows for the coherent superposition of possibilities, which demonstrate different behaviours depending on the *phase* between them. These behaviours constitute as *interference phenomena*, and lie behind the existence of algorithms in quantum computing which are arguably faster than the best classical alternatives. The concept of phase is deeply steeped in the structure of Hilbert spaces: the mathematical framework that underlies quantum theory. What if quantum theory did not hold in all scenarios, or was only a limiting case of some broader theory? In this case, would we still be able to meaningfully talk about phase and interference?

In this thesis, we will adopt an operational generalisation of quantum theory known as the *framework of generalised probabilistic theories*. We will provide a reasonable definition of phase in this framework. Using this, we shall explore single-particle interferometry set-ups (particularly Mach-Zehnder interferometers): experiments whose output is highly dependent on the phase between the spatially disjoint branches through which a particle might be traversing. By applying physically-motivated locality considerations, we identify the crucial role that the uncertainty principle and its generalisations play in quantum theory as an enabler of non-trivial interference. By taking into account relativity of simultaneity, we will also provide a physical motivation for why standard quantum theory provides the best description of the location of a particle traversing such a system. Finally, we apply our generalised definition of phase in the related context of particle exchange behaviour, and identify a method for classifying post-quantum particles.

All of this will demonstrate that phase between possibilities and its consequences are not uniquely quantum phenomena. Much of the behaviour we might ascribe to phase in quantum theory in fact holds generally true for phase in probabilistic theories.

Acknowledgements

I thank the Engineering and Physical Sciences Research Council (EPSRC) for funding my studentship, and am also grateful for the generous hosting provided to me by the *Centre for Quantum Technologies* at the *National University of Singapore* on occasions during the course of my graduate studies.

My warm personal gratitude goes to everyone I have collaborated with, including: Cormac Browne, Mile Gu, Markus Müller, Mio Murao, Yoshifumi Nakata, Jayne Thompson, and Nicole Yunger Halpern. I am also extremely grateful for frequently illuminating conversations with Felix Binder, Ross Dörner, Tristan Farrow, John Goold, Nana Liu, Chiara Marletto, Kavan Modi, Wonmin Son, Benjamin Yadin, and especially with Felix Pollock.

I am especially grateful for the keen eyes of Felix Binder, Tom Clucas, Oscar Dahlsten, David Garner, Nana Liu, and Benjamin Yadin, who helped with the proof-reading of this text at very short notice. Any errors which remain are mine, not theirs!

Extra special thanks goes to Oscar Dahlsten, who introduced me to the world of GPTs, and provided invaluable guidance and feedback throughout my graduate studies. Once, I spotted an advert in an airport which promised “strength, intelligence, and good looks”, and I commented that this covered all the qualities that anyone could want– but Oscar corrected me, “What about kindness?” He was right; for it is his kindness out of all his qualities for which I am the most grateful.

Finally, I thank my supervisor Vlatko Vedral for introducing me into the field of quantum information theory, and for regular discussions that often left me with a clearer understanding of physics, and always enthused me about my work; as well as for showing me how to appreciate the finer aspects of the academic lifestyle in Oxford and Singapore.

—Andrew J. P. Garner
19th January 2015.

List of publications

The text and graphics as presented in this thesis are entirely the original work of the author. Some of the scientific concepts and variations of the results have appeared in publications involving the author in collaboration with others. The relevant publications represented in this thesis are:

1. [1] A. J. P. Garner, O. C. O. Dahlsten, Y. Nakata, M. Murao, and V. Vedral (2013). *A framework for phase and interference in generalized probabilistic theories*. New Journal of Physics, **15** (9) 093044. doi:10.1088/1367-2630/15/9/093044. [Featured in chapter 3 and chapter 4.]
2. [2] O. C. O. Dahlsten, A. J. P. Garner, and V. Vedral (2014). *The uncertainty principle enables non-classical dynamics in an interferometer*. Nature Communications, **5** (4592). doi:10.1038/ncomms5592. [Featured in chapter 5.]
3. [3] A. J. P. Garner, M. P. Müller, and O. C. O. Dahlsten (2014). *The quantum bit from relativity of simultaneity on an interferometer*. (pre-print) arXiv:1412.7112 [quant-ph]. [Featured in chapter 6.]
4. [4] O. C. O. Dahlsten, A. J. P. Garner, J. Thompson, M. Gu, and V. Vedral (2013). *Particle exchange in post-quantum theories*. (pre-print) arXiv:1307.2529 [quant-ph]. [Featured in chapter 7.]

During the course of his graduate studies in Oxford, the author also contributed to the following papers on thermodynamics, unrelated to the central theme of this thesis:

5. C. Browne, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral (2014). *Guaranteed Energy-Efficient Bit Reset in Finite Time*. Physical Review Letters, **113** (10) 100603. doi:10.1103/PhysRevLett.113.100603.
6. N. Yunger Halpern, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral (2014). *Unification of fluctuation theorems and one-shot statistical mechanics*. (pre-print) arxiv:1409.3878 [cond-mat].

Contents

1	Motivation and general overview	1
1.1	Physics beyond quantum theory	1
1.2	What is phase?	2
1.3	Overview of results	4
2	Introduction to the GPT framework	5
2.1	Classical probability theory	6
2.2	Quantum systems	9
2.2.1	Quantum theory for pure states	9
2.2.2	Mixed states of quantum systems	13
2.2.3	Quantum measurements	15
2.3	The framework of generalised probabilistic theories (GPTs)	17
2.3.1	The generalised experiment	18
2.3.2	States	19
2.3.3	State-spaces	20
2.3.4	Effects	22
2.3.5	Transformations	23
2.3.6	Theories	25
2.4	Example theories within the GPT framework	25
2.4.1	Classical probability theory: <i>bits</i>	25
2.4.2	Quantum theory: <i>qubits</i>	27
2.4.3	Beyond quantum theory: <i>gbits</i>	29
2.5	Minimum-dimension representation	31
2.6	Multi-partite systems in the GPT framework	32
2.6.1	The tensor product for states and for effects	34
2.6.2	Example: P-R boxes	35
2.7	Spekkens' toy model	37
2.7.1	Transformations	39
2.7.2	Multi-partite systems	40
2.7.3	Convex extension	43
3	Phase in the GPT framework	46
3.1	Phase in quantum systems	47
3.1.1	Phase in pure quantum states	47
3.1.2	Phase in general quantum states	48

3.2	Defining phase in the GPT framework	49
3.2.1	The structure of phase transformations	50
3.3	Examples of phase groups	53
3.3.1	Classical probability	53
3.3.2	Quantum theory: qubits	54
3.3.3	Box-world: 3-in 2-out gbits	55
3.3.4	Spekkens' toy model	57
3.4	Examples of irreversible phase dynamics	59
3.4.1	Decoherence	59
3.4.2	'Measurement setting' in a gbit	60
3.5	Phase transformations of <i>maximal measurements</i>	61
3.5.1	Phase and non-classicality	63
3.6	The phase dynamics of related measurements	65
3.7	POVMs and general measurements	68
3.7.1	Unambiguous quantum state discrimination	69
3.7.2	Informationally complete measurements	71
3.8	Conclusion	72
4	Interference in the GPT framework	73
4.1	The quantum Mach–Zehnder interferometer	73
4.2	Mach–Zehnder interferometers in the GPT framework	76
4.2.1	Example: Interference for 3-in 2-out gbits	78
4.2.2	Example: Interference in Spekkens' toy model	79
4.3	Discussion of interference in the GPT framework	81
4.3.1	Further abstraction of the Mach–Zehnder interferometer	81
4.3.2	Other treatments of interference in the GPT framework	82
4.4	Conclusion	83
5	Interference, locality, and uncertainty	84
5.1	Branch locality	85
5.2	Branch locality in quantum theory	88
5.2.1	Reversible quantum transformations and branch locality	88
5.2.2	Irreversible quantum transformations and branch locality	90
5.2.3	Example branch-local operation: decoherence	93
5.2.4	Non-branch local quantum operations	94
5.2.5	The Aharanov–Bohm effect	95
5.3	Fully independent systems	95
5.3.1	Example: 2-in 2-out gbit	96
5.3.2	Fully independent systems in general	97
5.4	Fully conditionally-restricted systems	101

5.5	Quantum-like uncertainty relations	103
5.6	Mutually unbiased measurements	105
5.7	Trade-off between dynamical and state-space freedoms	109
5.7.1	Example: Qubit interferometer	110
5.7.2	Example: Stabilizer interferometer	111
5.7.3	Example: ‘Square-based pyramid bit’ interferometer	112
5.7.4	Example: Spekkens’ toy model interferometer	113
5.8	Deutsch’s algorithm beyond quantum theory	115
5.9	Extensions to branch locality	117
5.10	Conclusion	120
6	Special relativity and interferometers	122
6.1	Relativity and the GPT framework	123
6.2	Relativistic behaviour of interference experiments	124
6.3	Simulating interferometers	128
6.3.1	Remark: Random access Deutsch’s algorithm	132
6.4	n -balls and the dimensionality of quantum theory	132
6.4.1	Subgroups of the orthogonal group	134
6.4.2	Interpretation of the four-dimensional case and synopsis	137
6.5	Conclusion and outlook	139
7	Post-quantum particle exchange	140
7.1	Exchanging subsystems in quantum theory	140
7.1.1	Indistinguishable quantum states	142
7.2	Swapping and indistinguishability in post-quantum theories	144
7.2.1	Locally tomographic theories, and local distinguishability	145
7.2.2	Post-quantum controlled-swap	147
7.3	Generalised particle types	148
7.3.1	Example: classical theory	149
7.3.2	Example: ball state-spaces	150
7.3.3	Example: ‘box-world’	151
7.3.4	Example: Spekkens’ toy model	153
7.4	Conclusion and outlook	154
8	Conclusion	155
A	Uncertainty in quantum theory	157
B	Non-existence of controlled-swap in Spekkens’ toy model	159
	Bibliography	165

Chapter 1

Motivation and general overview

1.1 Physics beyond quantum theory

Quantum theory has been used with phenomenal success as a tool in the laboratory to predict the behaviour of systems observed in experiments. It has also been used to provide explanations of complex behaviours in terms of simpler quantum concepts. However, the *foundations of quantum theory* have been notoriously difficult to interpret or motivate, and thus the study and discussion of them forms a broad field¹ of academic debate, both in physics and in philosophy.

As tempting as it would be to adopt the “shut up and calculate” [5] approach (often attributed to the *Copenhagen interpretation* of quantum mechanics [6]), this is a very short-sighted position to take. Throughout the history of physics, seemingly reasonable theories have been replaced by newer alternatives where the prior theory is just a limiting case (e.g. special and general relativity superseding Newtonian mechanics). One would have to either be incredibly bold or stubbornly unimaginative to claim to have finally found the ultimate description of reality. In hindsight, statements such as those allegedly made by in 1900 by Lord Kelvin² that “*There is nothing new to be discovered in physics now; all that remains is more and more precise measurement,*” [7] have proved to be very wrong indeed.

In order to even conceive of how quantum theory could change as science progresses, we cannot start from inflexible mathematical statements such as “the state of a system can be described using a complex Hilbert space.” Instead, it is better to motivate *why* we should adopt Hilbert spaces (see e.g. [8]) in the first place, by the combination of fundamental *axioms*: reasonable statements about reality that we believe are true. Not only will such statements provide a more satisfying explanation of why quantum theory is as it is, a well-chosen axiom will also provide us with rules that might be experimentally tested in the search for alternative theories, and falsifiable alternatives should it not in fact be true.

¹ See, for instance, the journal *History & Philosophical Foundations of Physics*, Springer.

²According to [7], a similar statement may have also been made by Michelson around the same time.

This methodology is known as the *axiomatic approach* to quantum theory. There have been many attempts at this including (amongst many other notable examples) those presented in [9–17]. In this thesis, we shall adopt the framework of generalised probabilistic theories (see for example [9, 10, 14, 18], and chapter 2 of this text), which provides a mathematical structure allowing us to compare quantum theory with alternative theories on an even footing. This framework is *operational* in that it is concerned with observable statistics (rather than the underlying theory that generates them) and so allows for the imposition of an axiomatic structure on top of it (as was done in [9]), which can be used to select specific descriptions of reality, such as quantum and classical theory.

Although the goal of this thesis is not to provide a complete axiomatic derivation of quantum theory, the work here is undertaken in a similar spirit. We shall impose reasonable behaviours (e.g. axioms regarding locality, or adherence to the rules of special relativity), and see if this causes deviation in behaviour from that predicted by quantum mechanics in the specific context of *phase*.

1.2 What is phase?

The concept of phase has long-existed in physics. It pre-dates quantum theory, as its origins lie in classical wave-mechanics, where it plays the mathematical role of describing the current stage of a cyclic process by a real number. For instance in the periodic solution to the wave equation $Y = A \cos(\omega t)$, every value ωt , $\omega t + 2\pi$, $\omega t + 4\pi$ corresponds to the same value of Y . Here the value $\phi = \omega t \bmod 2\pi$ gives the stage of the wave-motion in terms of a number between 0 and 2π . This value is known as the *phase* of the wave. This mathematical role of phase appears in the behaviour of classical waves, including those of water waves and electromagnetic radiation.

Phase is responsible for the phenomenon of *interference*, which occurs when two or more waves combine. In general, to predict the behaviour of two combining waves, it is not sufficient just to know the amplitudes (in the example, A) of the waves. If the waves are *in phase* (such that ϕ agrees between the two cases up to a multiple of 2π), then one has *constructive interference* where the net effect is that of a single wave of the sum of amplitudes [e.g. $Y = A_1 \cos \phi + A_2 \cos \phi = (A_1 + A_2) \cos \phi$]. On the other hand, if the waves are *anti-phase* such that ϕ differs by π , then one has *destructive interference* where the net effect is a single wave of the difference of amplitudes, or even complete cancellation if the amplitudes are the same [e.g. $Y = A_1 \cos \phi + A_2 \cos(\phi + \pi) = (A_1 - A_2) \cos \phi$, which is 0 if $A_1 = A_2$]. This vastly different behaviour, dependent on whether two waves are in or out of phase, highlights the importance of knowing the phase as well as the amplitude of a wave in order to predict

how it interacts with other waves.

In this thesis, we are specifically interested in phase as it appears in quantum theory. The role played by phase in quantum theory is mathematically similar to that in classical wave mechanics. However, rather than waves relating to the displacement of water, or the strength of an electric field, in quantum theory there are waves of *probability*, arising as the solutions to the Schrödinger equation (see e.g. [19]). This allows not just for the superposition of possibilities, but also the cancellation of possibilities. Should a system admit description of its location by the superposition of two wave functions $\psi_1(x)$ and $\psi_2(x)$ that are in anti-phase at x , then there will be no probability of detecting that system at x ; even if $\psi_1(x)$ and $\psi_2(x)$ both independently have some probability of being at x . This is known as *coherent* behaviour: if the probabilities had just been *mixed together*, such as by the laws of classical probability $P(X) = \sum_i P(\psi_i) P(X|\psi_i)$ (where the system is described by ψ_i with probability $P(\psi_i)$, $P(X)$ is the probability of finding the system at X , and $P(X|\psi_i)$ is the probability of finding the system at X given it is in ψ_i), then this cancellation would not occur.

This wave-like behaviour carries through when we abstract the description of the system into the Hilbert spaces popularly used in quantum information theory (see, e.g. [8], and discussion in § 2.2 and § 3.1). To fully describe a quantum state of a system, not only must we associate a probabilistic weighting for how likely each possible configuration is, we must also associate a *phase* that encodes how it might interfere with the other possibilities.

This interaction between possibilities yields some of the most exciting quantum phenomena. An extreme example lies in the famous thought-experiment of Schrödinger's cat [20], where there is a crucial difference between the quantum and classical cat. For the quantum cat, whose health is in a *superposition* of dead and alive, these coherent alternatives can be made to interfere with each other in order to return the cat to a state of being wholly alive. On the other hand, for a classical cat (who might also have a 50% chance of being alive), there will be no way to undo what might have happened to the poor creature. Similar behaviour accounts for the arguable speed-up of quantum computers over classical alternatives [21–27]. For a quantum computer, one can input a superposition of every plausible classical bit string (sequence of 0s and 1s) *at the same time*. The quantum computer then interferes these alternative inputs to produce some useful result about the problem under consideration. Attempting to run such an algorithm with a classically mixed input of bit strings would only yield a meaningless random result.

Thus, we consider it worthy to establish whether the concept of phase transfers to post-quantum theories, and how much of the related phenomena can be recovered in general probabilistic theories.

1.3 Overview of results

We shall proceed in chapter 2 with a technical introduction. We review how the state of a system may be described in classical and quantum theory, and from this perspective introduce the framework of generalised probabilistic theories (GPT framework). We shall also provide some details about Spekkens’ toy model [28, 29], a foil for quantum theory, which has a natural adaptation into the GPT framework.

In chapter 3, we present the first novel results of this work: a general definition of phase in the GPT framework. We show how this definition is consistent with the quantum case, and identify instances of phase in other probabilistic theories. We will prove how this definition is intrinsically related to ‘non-classicality’, and show how it can be related to decoherence, as well as applied in the context of general quantum measurements (POVMs).

In chapter 4, we apply the ideas in the preceding chapter to the physical context of a Mach–Zehnder interferometer, and demonstrate how interference patterns might be observed in non-quantum theories.

In chapter 5 we introduce the concept of *branch locality*, which identifies what consequences can be induced by independent agents acting locally on the spatially-disjoint branches of an interferometer. We show that this could in general place severe restrictions on the interference observable, but that this effect is mitigated in quantum theory by the uncertainty principle. We will consider the Deutsch-Jozsa algorithm [21, 22], a protocol for *quantum computing*, that solves a specific problem faster than any known deterministic classical algorithm, and see which non-quantum theories have the ability to execute this protocol using the Mach–Zehnder interferometer.

In chapter 6, we will also consider the consequence of *relativity of simultaneity* (a concept from special relativity) [30], as applied to a single particle travelling through the Mach–Zehnder interferometer, and demonstrate that this puts further restrictions on physically reasonable descriptions of the state. In particular, we consider a family of theories known as *n*-balls (e.g. [14]), and see that the relativistic consideration singles out quantum dynamics (i.e. the three-dimensional Bloch sphere).

Finally, in chapter 7, we consider another application of phase in quantum theory: the behaviour of indistinguishable particles under exchange. We introduce an operational definition of indistinguishability, and show how a quantum experiment that identifies different particle types (e.g. bosons/fermions) can be extended into the GPT framework. We consider how some of these post-quantum particles might behave.

All of this, taken together, shall demonstrate that phase between possibilities and its consequences are not uniquely quantum phenomena. We will see that much of the behaviour that we might have ascribed to quantum theory in fact holds true for probabilistic theories in general.

Chapter 2

Introduction to the GPT framework

The essence of science is the elegant description of reality. Mathematical science concerns itself with finding mathematical constructions that allow quantifiable predictions about the behaviour of a physical system. For this to be possible, there must be a way of assigning a mathematical object to the state of the system, as well as to the various mechanisms by which the physical system evolves.

In this thesis, we shall be particularly concerned with *operational* descriptions of reality, in which the state of the system is described by a set of statistics, and the evolution of a system is represented by maps that transform these statistics into new statistics. In such a framework, physical laws become statistical restrictions. The use of such an approach should not be mistaken for adopting an instrumentalist philosophy (which holds that the *only* valid science is the ability to make falsifiable predictions); this framework can also be used to provide *explanations* within operational descriptions: ways of connecting otherwise a priori unrelated physical laws to observable phenomena via their observable statistics (see especially chapter 5 and chapter 6).

In this chapter, we shall outline this operational approach. We shall begin by considering how a classical system might be described by probabilities (§ 2.1). We shall then proceed through a lightning introduction to the description of a system using quantum theory, and how this differs from the classical case (§ 2.2). We subsequently identify shared features of these two cases, and show how they may be abstracted into a framework of generalised probabilistic theories (GPTs) (§ 2.3). We will provide several examples (§ 2.4), introduce an alternative representation of states (§ 2.5), and consider in particular the behaviour of multiple systems within the framework (§ 2.6). In the concluding section of this chapter (§ 2.7), we will introduce Spekkens' toy model: a thought experiment that is classical in nature but shares many properties with quantum theory. Although this model is not a priori within the GPT framework, we shall discuss how it can be extended such that it may be treated as if it is.

2.1 Classical probability theory

Classical probability theory provides the simplest model for the description of any system. A classical system admits a single largest set of mutually exclusive possibilities¹ (e.g. the numbers 1 to 6 on a dice; “heads” and “tails” on a coin; or the values 0 and 1 in a bit).

When the classical system is known to be in one of these possibilities with certainty (and thus to the exclusion of all other possibilities), then the system is said to be in a **pure state**. However, pure states alone are not sufficient to describe every possible state of knowledge about the system. Consider an experiment where one trial consists of flipping a fair coin and hiding the result. If the observer assigned either heads or tails to the state of the system for approximately half of the trials their description would be correct, and the other half, their description would be incorrect.

Probability theory allows for a description of the state of the system that reflects this uncertainty. One can describe the state of the object in terms of the relative frequency² with which it is to be found in one possibility over the others, in the limit of an infinite number of repeated trials. For every possible outcome i , one may assign a real number p_i in the range $[0, 1]$; and provided that it is guaranteed there will be an outcome at all (such that the system is said to be *normalised*), then $\sum_i p_i = 1$. This set of assignments of probabilities to possibilities may be written as a real vector of dimension equal to the number of possibilities. Such a vector is known as the **probability distribution** or **probability vector**.

For instance, for a system that can be in one of N different possibilities labelled $\{z_1, \dots, z_N\}$, then an assignment of probabilities to possibilities s would be written:

$$s = \begin{pmatrix} P(z_1) \\ \vdots \\ P(z_N) \end{pmatrix} \quad (2.1)$$

where $P(z_i)$ is the probability that the system is in z_i .

A particular choice of the vector s is known as the **state** of a system. In general, a state is an epistemic property in that it reflects the knowledge of an agent about the

¹Possibilities may be called *branches* if the freedom is to do with which branch of an interferometer a particle is in; or *levels* if the possibilities are associated with a system having different energies. One might also call such a possibility a state, but we will avoid using this term in the context of possibilities.

²Discussing the philosophical nature of probability is not the purpose of this thesis (see instead [31] and references within). Those who are dissatisfied with the relative frequency outlook on probability may view it as a purely mathematical weighting associated with a possibility given to options in rational betting strategies (propensity).

system. For example, consider two agents playing a game where a fair coin is flipped and hidden. One agent may look at the coin and see it to be definitely in a state of heads $(1, 0)^T$, whereas the other agent who does not have this knowledge (but knew that the coin being flipped was fair) would assign it the state $(\frac{1}{2}, \frac{1}{2})^T$. Regardless of this mixed distribution, if the second agent were to look at the coin, she would definitely find the coin to be in a state of heads. It may be assumed classically that the system has an ontic state (reflecting *reality*) which is always pure, regardless of what an agent says about it. This realism is not so clear-cut for systems described by quantum and post-quantum theories.

A transformation on a classical state is a map that changes the assignment of probabilities in some way. The output of any transformations must itself be a valid probability vector. When we assume that the updated state depends *only* on the current state, the transformation is *Markovian*. In particular, the set of Markovian linear maps (linearity will be motivated in § 2.3.5) correspond exactly to the left-multiplication of probability vectors by (left-)stochastic matrices (see e.g. [32]): $N \times N$ non-negative matrices, in which each column sums to 1. If there are other restrictions, then the set of allowed transformations will be a subset of these.

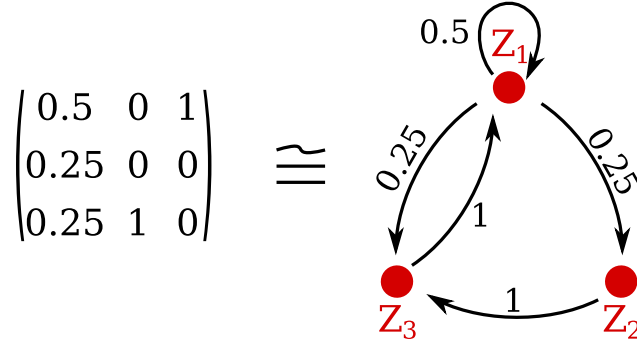


Figure 2.1: Directed graph of a 3×3 stochastic matrix For a stochastic matrix (LHS) acting on a system with three possibilities z_1 , z_2 and z_3 , with states given by the vector $s = [P(z_1), P(z_2), P(z_3)]^T$, we can draw the directed graph on the RHS. The red nodes correspond to the different possibilities, and the arrows from each node correspond to a potential transition from one possibility to another, with a probability given by the associated numerical value.

We can use the isomorphism between non-negative matrices and directed-graphs [33] to represent stochastic matrices by a network diagram (e.g. fig. 2.1), where the nodes are the possibilities, and the directed-lines are labelled with the conditional probability of transforming the state from $z_j \rightarrow z_i$, given by the element M_{ij} of a stochastic matrix M .

Reversible transformations correspond to permutations on the labels of the measurement outcomes, and so are represented by the permutation group S_N . Considering

only pure states (where the system is in one possibility with certainty) which could be treated as ontic states, a permutation could be thought of as the map consisting of a series of physically implementable instructions “if the system is in z_i , move it to $\sigma(z_i)$ ” where σ maps each pure state onto a possibly different pure state. Similarly, the evolution of an underlying ontic state for an arbitrary stochastic matrix may be considered as a random walk on the associated network diagram.

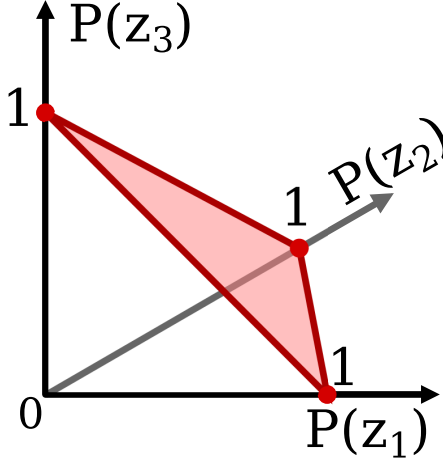


Figure 2.2: Classical simplex state-space. The state-space of a three-level classical system is an equilateral triangle (2-simplex). The three normalised pure states (red dots) lie on the three axes, equidistant from each other. The red plane shows the set of all valid normalised states.

Finally, we remark that (in the absence of any other restrictions on the allowed states) the set of all allowed states (known as the **state-space**) of a classical system is given by a regular simplex: the generalisation of point, line, triangle, tetrahedron into arbitrary dimensions. This simplex structure can be seen by drawing the probability vector (eq. (2.1)) in Euclidean space, such that the probability of each outcome corresponds to a different Cartesian axis (see fig. 2.2). If we consider only the states where $\sum_i P(z_i) = 1$ (which is the equation for a hyperplane of dimension $N - 1$ whose normal is given by $\mathbf{1} = (1, \dots, 1)$), we note that each axis only contains one point, corresponding to one $P(z_i) = 1$ and all other $P(z_j) = 0$. These are the pure states of the system, and each is exactly the same distance from all the other pure states. This corresponds to the geometric definition of a regular simplex: a shape whose vertices are mutually equidistant. We note that the composition (chaining) of permutations of these vertices gives the automorphism group of the simplex S_N , and there is a one-to-one mapping of every state in the space.

A measurement in classical theory consists of checking which possibility the object is in. When an agent measures a state $\mathbf{s} = (P(x_1), \dots, P(x_n))^T$, she expects to find the system to be in possibility x_i with probability $P(x_i)$, and if this is indeed the outcome

she gets, she would update her state to $\mathbf{s}' = (0, \dots, 0, 1, 0, \dots)^T$ which only has support on the element $P(x_i)$. If a measurement was made on a system previous in state $\mathbf{s}$, but the agent either does not learn the outcome or forgets the outcome, then the state she should assign to the system afterwards is still $\mathbf{s}$.

2.2 Quantum systems

2.2.1 Quantum theory for pure states

Now let us consider describing a system with N mutually exclusive possibilities by quantum theory, rather than by classical probability theory. Such a quantum system is qualitatively different from an N -level classical system: in the classical case, knowing the probability of each of these N possibilities is sufficient to determine everything there is to know about the system (i.e. its state), but this is not the case for a quantum system. This is because quantum theory also admits *superpositions* of possibilities: descriptions of the system in which none of the possibilities occur with certainty, but the state of the system is still considered in some way to be pure.

It is almost the definition of quantum system that its state is described by a complex Hilbert space. A full pedagogical introduction to Hilbert spaces may be found (for instance) in chapter 2 of [8], but here we will briefly outline the machinery we require. In essence, a complex Hilbert space consists of a complex vector space, with an inner product $I(\cdot, \cdot)$ that maps two vectors onto a single scalar complex number, according to the rules: **i.** for any two vectors $|\phi\rangle$ and $|\xi\rangle$ (written in the Dirac notation [34, 35]), then $I(|\phi\rangle, |\xi\rangle) = I(|\xi\rangle, |\phi\rangle)^*$ where $*$ represents complex conjugation; **ii.** Linearity in one argument³ $I(|\phi\rangle, c_1|\xi_1\rangle + c_2|\xi_2\rangle) = c_1I(|\phi\rangle, |\xi_1\rangle) + c_2I(|\phi\rangle, |\xi_2\rangle)$ for vectors $|\phi\rangle, |\xi_1\rangle, |\xi_2\rangle$ and complex scalars c_1 and c_2 ; **iii.** $I(|\phi\rangle, |\phi\rangle) \geq 0$ for all $|\phi\rangle$, and equality holds if and only if $|\phi\rangle$ is zero.

For a space of vectors (“kets”) $\{|\phi\rangle\}$, we can form a dual space of covectors (“bras”) $\{\langle\phi|\}$ which act as functions whose argument is a ket, and whose output is a single complex number. In fact, from the Fréchet-Riesz theorem (see e.g. §13.3 in [36]), the set of bras form an isomorphic Hilbert space to the kets, with a related inner product function defined as $\bar{I}(\cdot, \cdot)$ satisfying $\bar{I}(\langle\phi|, \langle\xi|) = I(|\xi\rangle, |\phi\rangle) = I(|\phi\rangle, |\xi\rangle)^*$. This one-to-one correspondence, taken with the fact that a bra can map a ket to a complex number allows us to express the inner product between two kets as the application of an associated bra onto a ket : $I(|\phi\rangle, |\xi\rangle) = \langle\phi|\xi\rangle$.

³i and ii together imply anti-linearity in the other argument.

The kets of a Hilbert space can either represent finite-length vectors, or continuous functions (such as a solution to the Schrödinger equation), which may be casually thought of as infinite length vectors where element x is given by the value of the function $\phi(x)$. The bra-ket notation can be used in either case. For finite vectors $|\phi\rangle = (\phi_1, \dots, \phi_N)^T$ and $|\xi\rangle = (\xi_1, \dots, \xi_N)^T$, the inner product is explicitly the element-wise sum $\langle\phi|\xi\rangle = \sum_{i=1}^N \phi_i^* \xi_i$. For a choice of continuous functions $\phi(x) \rightarrow |\phi\rangle$ and $\xi(x) \rightarrow |\xi\rangle$, the inner product is equivalently given by the integral $\langle\phi|\xi\rangle = \int_{-\infty}^{\infty} dx \phi^*(x) \xi(x)$.

Let us return to the description of the state of a system. If a system has N mutually exclusive possibilities, the quantum description is via an N -dimensional complex Hilbert space. N complex numbers naturally admit more freedom of choice than the N real numbers of a classical probability vector, and we see that using a Hilbert space description gives rise to the concept of additional **complementary measurements** that do not exist in the equivalent classical description of the state.

Consider a pure state of a **qubit** (two-level quantum system), written as a two-element complex vector:

$$|\phi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha|0\rangle + \beta|1\rangle \quad \alpha, \beta \in \mathbb{C}, \quad |\alpha|^2 + |\beta|^2 = 1 \quad (2.2)$$

where $|0\rangle = (1, 0)^T$ and $|1\rangle = (0, 1)^T$ are two vectors forming an orthonormal basis (their inner product satisfies $\langle 0|1\rangle = 0$, and $\langle 0|0\rangle = \langle 1|1\rangle = 1$).

Each element of the complex vector is associated with the probability of finding the system in a particular possibility. Explicitly, according to the *Born rule* [37], the probability of finding a system described by $|\phi\rangle$ to be in the level labelled by 0 is given by $|\langle 0|\phi\rangle|^2 = |\alpha|^2$, where $\langle 0| = (1, 0)$ is the dual of the (unique) normalised state associated with being definitely in level 0. Likewise the probability of being in 1 is given by $|\langle 1|\phi\rangle|^2 = |\beta|^2$. Following the measurement, the standard quantum update rule (von Neumann projection in this case, Lüder's rule in general; see e.g. [38]) is that the system is now in state $|0\rangle$ if the outcome was 0, or $|1\rangle$ if it was 1.

However, under the Born rule, all states of the form $|\phi\rangle = a|0\rangle + be^{i\theta}|1\rangle$ for $a, b \in \mathbb{R}$ such that $a^2 + b^2 = 1$, and $\theta \in [0, 2\pi)$ will give the same probabilities of being in 0 (or in 1), regardless of the choice of θ . The state hence encodes some additional information beyond what can be probed by a measurement of possibility 0 against possibility 1. For this extra freedom to have an operational meaning, there must be some way of observing a consequence if θ were to differ between two states.

The term in θ becomes important when we consider the allowed quantum transformations. Quantum theory admits a more general class of reversible transformation than just permuting elements of the vector. The reversible transformations of an N -level quantum system correspond to pre-multiplying the ket by an $N \times N$ unitary matrix. The unitary matrices are defined such that they must preserve the Hilbert space inner

product $I(U|\psi\rangle, U|\xi\rangle) = \langle\psi|U^\dagger U|\xi\rangle = \langle\psi|\xi\rangle$ for any choice of vectors $|\xi\rangle$ and $|\psi\rangle$. This implies that they satisfy $U^\dagger U = \mathbb{1}$, and hence $U^\dagger = U^{-1}$ (making it always possible to find a unique transformation that reverses the effect a unitary has on any system). Unlike the classical permutation group S_N which has a finite number of elements, $U(N)$ is a compact Lie group (see e.g. [39, 40]) which has an infinite number of elements.

One possible transformation on a qubit would be the following unitary transformation (specifically known as a *Hadamard gate*):

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \in U(2), \quad (2.3)$$

which evolves the state $|\phi\rangle = a|0\rangle + be^{i\theta}|1\rangle$ into $|\phi'\rangle = H|\phi\rangle = (a + be^{i\theta})|0\rangle + (a - be^{i\theta})|1\rangle$. The probability of measuring 0 or 1 in the new state $|\phi'\rangle$ is given by:

$$P(0|\phi') = |\langle 0|H|\phi\rangle|^2 = \frac{1}{2} + ab \cos \theta, \quad P(1|\phi') = |\langle 1|H|\phi\rangle|^2 = \frac{1}{2} - ab \cos \theta, \quad (2.4)$$

and hence the initial choice of θ is important to predict the “which possibility?” measurement in $|\phi'\rangle$, and thus has operational meaning.

As unitary operators do not change the inner product in a Hilbert space, $I(|x\rangle, U|y\rangle)$ is equivalent to $I(U^\dagger|x\rangle, U^\dagger U|y\rangle) = I(U^\dagger|x\rangle, |y\rangle)$, and $(U^\dagger|x\rangle)^\dagger = \langle x|U$. This allows us to unambiguously write the expression $\langle x|U|y\rangle$ where U could either be a pre-multiplication on $|y\rangle$ or a post-multiplication on $\langle x|$ with the same (complex scalar) result. Thus it is mathematically equivalent to treat the role H plays in the expressions eq. (2.4) as acting not to change $|\phi\rangle \rightarrow |\phi'\rangle$, but rather to transform $\langle 0|$ and $\langle 1|$ to $\langle +| \equiv \langle 0|H = \frac{1}{\sqrt{2}}(\langle 0| + \langle 1|)$ and $\langle -| \equiv \langle 1|H = \frac{1}{\sqrt{2}}(\langle 0| - \langle 1|)$ respectively. Mathematically, $|\langle +|\phi\rangle|^2$ and $|\langle -|\phi\rangle|^2$ recover *exactly* the same probabilities as in eq. (2.4). Thus, we can think of the ‘transform and measure’ operation described above as in fact just being an alternative *complementary* measurement of the system in a new basis which compares the ‘ $|+\rangle$ -ness’ vs. the ‘ $|-\rangle$ -ness’ of the state $|\phi\rangle$. In this basis, the state $|\psi\rangle$ could be written

$$|\psi\rangle = \frac{1}{\sqrt{2}} (a + be^{i\theta})|+\rangle + \frac{1}{\sqrt{2}} (a - be^{i\theta})|-\rangle \quad (2.5)$$

and this object represents exactly the same physical state of the system as $|\phi\rangle = a|0\rangle + be^{i\theta}|1\rangle$.

In an abstract information-theoretical context, there is little distinction between the first approach of actively transforming a state into a new state and measuring in a fixed basis, and the second approach of passively transforming the basis so that the state is in a new co-ordinate system. It is implicitly assumed that there is some way of measuring the system in whatever basis we choose. However, in physical realisations of measurement, it may be that making the measurement in different bases does in fact

involve an active transformation on the state. Consider the set-up in fig. 2.3 where a photon travels down one of two channels, with the “which path?” degree of freedom described by a qubit. The photon detectors at the end of each channel click whenever a photon is present, and so naturally measure in the $|0\rangle$ vs $|1\rangle$ “which path?” basis. To measure the full quantum state (a process known as *tomography*) requires the insertion of a coherent quantum beam-splitter that implements a transformation on the state (such as the unitary H above) before it enters the detectors.

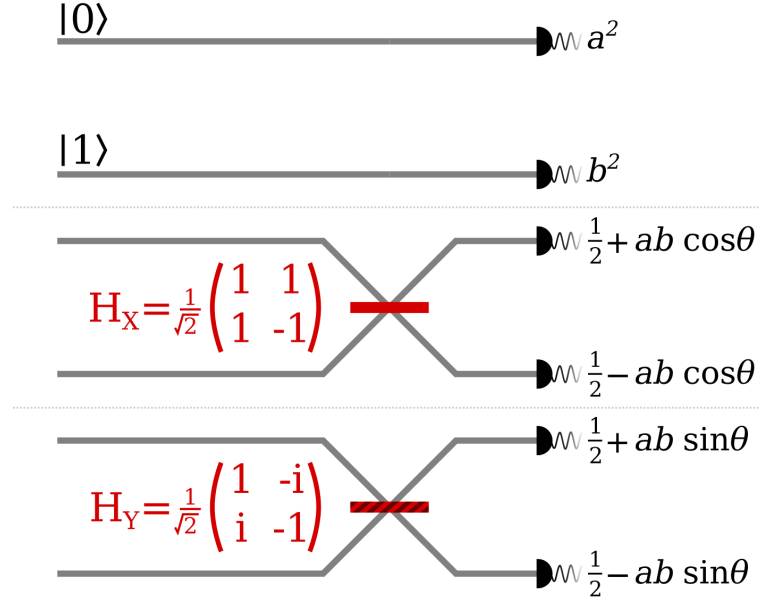


Figure 2.3: Measuring a qubit state. A photon travelling down one of two branches is described by the quantum state $|\phi\rangle = a|0\rangle + be^{i\theta}|1\rangle$. The choice of beam-splitter inserted (corresponding to a choice of a physical optical component), affects the statistics of the output measurement. The expressions on the right hand side reflect the probability of a click in that particular beam-splitter given the state started in $|\phi\rangle$. In general, for a qubit three choices (no beam-splitter whatsoever, such as in the top diagram, is a valid choice) are required to completely determine the state, as there are multiple values of $\theta \in [0, 2\pi)$ corresponding to the same value of $\cos \theta$.

In fact, for every complex vector in quantum theory, it is possible to come up with a unitary matrix that allows the system to be described as a single supported possibility of *some* measurement. It is this reason why quantum systems in a state described by a complex vector are called pure, even if for most measurements the outcome cannot be predicted with certainty.

2.2.2 Mixed states of quantum systems

Although the pure state of an N level quantum system can be written as a complex vector of the form $|\psi\rangle = \sum_i c_i |i\rangle$ for $\{c_i\} \in \mathbb{C}$ and a choice of mutually exclusive outcomes $\{|i\rangle\}$, states of this form do not encompass all possible states of knowledge that could be held about a quantum system. There are also **mixed states**, which mathematically can be thought of as the combination of classical probability with quantum states.

Consider the scenario where with probability p , a machine will prepare a qubit system such that it would be described by an agent by the quantum state $|0\rangle$, and with probability $1 - p$ by the state $|1\rangle$. The pure states which assign the correct possibilities to finding the system in $|0\rangle$ or $|1\rangle$ are of the form $|\psi\rangle = \sqrt{p}|0\rangle + e^{i\theta}\sqrt{1-p}|1\rangle$. We shall demonstrate, however, that none of these are a good choice for the agent to assign to the state output from the machine. Suppose (for clarity of the example) that $p = \frac{1}{2}$, so that the candidate output state is $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\theta}|1\rangle)$. The $|+\rangle$ vs $|-\rangle$ measurement (which could be thought of as acting on the state with the H in eq. (2.3)) has probability $\frac{1}{2}(1 + \cos \theta)$ associated with finding the system in $|+\rangle$, and $\frac{1}{2}(1 - \cos \theta)$ with $|-\rangle$. On the other hand, a $|i\rangle$ vs. $|-i\rangle$ measurement, associated with first transforming the state by $H_Y = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}$, has associated probabilities given by $\frac{1}{2}(1 + \sin \theta)$ and $\frac{1}{2}(1 - \sin \theta)$ respectively (as in fig. 2.3).

Now suppose the system in fact outputs the state $|0\rangle$. The statistics associated with the X measurement are both $\frac{1}{2}$ for $|+\rangle$ and for $|-\rangle$. Likewise, for the Y measurement one also recovers $\frac{1}{2}$ for $|i\rangle$ and for $|-i\rangle$. The same is also true if the state output was $|1\rangle$. By the laws of combining classical probabilities, we hence expect that if we made the X measurement without knowing which of these two probabilities, the outcome would be $|+\rangle$ with probability given by $pP(|+\rangle | 0 \text{ prepared}) + (1 - p)P(|+\rangle | 1 \text{ prepared})$, where $P(X|Y)$ is read as “the probability of outcome X given condition Y ”. Hence, every probability associated with either outcomes of the X measurement or the Y measurement should be $\frac{1}{2}$. On the other hand, no value of θ satisfies $\frac{1}{2}(1 + \cos \theta) = \frac{1}{2}(1 - \cos \theta) = \frac{1}{2}(1 + \sin \theta) = \frac{1}{2}(1 - \sin \theta) = \frac{1}{2}$. This means that no single choice of θ can be assigned to the state $|\psi\rangle$ to ensure that it predicts the correct statistics for any measurement made on the system. We thus have found an example of a quantum system that cannot be described by a pure quantum state.

Although one could specify the state of systems by a probability distribution over complex-valued kets, there is a specific quantum object that encodes this concisely. Recall that the probability of an outcome $\langle 0|$ from a pure quantum state $|\psi\rangle$ is given by $|\langle 0|\psi\rangle|^2 = \langle 0|\psi\rangle\langle\psi|0\rangle$. Suppose there were a set of pure states $|\phi_1\rangle, \dots, |\phi_n\rangle$ each prepared with probability $p_1, \dots, p_n$. Classical probability theory tells us then that the chance of measuring an outcome $\langle 0|$ from a random choice from this set is given by

$P(0) = \sum_i P(0|\phi_i) P(\phi_i)$ and hence $P(0) = \sum_i p_i \langle 0|\phi_i\rangle \langle \phi_i|0\rangle$. By linearity, we can take the sum inside such that $P(0) = \langle 0|(\sum_i p_i |\phi_i\rangle \langle \phi_i|)|0\rangle$. The same process can be done for any measurement outcome $\langle x|$, and for a given preparation the object inside the $\langle x|\bullet|x\rangle$ will be the same in each case. Thus this object, known as the **density operator**, is a good candidate to use for our operational description of a mixed quantum state.

Let us remark that the outer product of two kets $|\phi\rangle$ and $|\xi\rangle$ written as $|\phi\rangle\langle\xi|$ is a linear function that maps a vector $|\xi\rangle$ to $|\phi\rangle$, such that $(|\phi\rangle\langle\xi|)|\xi\rangle = |\phi\rangle$, and in general maps a vector $|\alpha\rangle$ to a vector parallel to $|\phi\rangle$ with (complex) length given by $\langle\xi|\alpha\rangle$ (i.e. $|\phi\rangle\langle\xi|\alpha\rangle = \langle\xi|\alpha\rangle \cdot |\phi\rangle$). For finite dimensions where $|\alpha\rangle = \sum_i a_i |i\rangle$ and $|\beta\rangle = \sum_j b_j |j\rangle$, this may explicitly be written as a matrix:

$$|\alpha\rangle\langle\beta| = \begin{pmatrix} a_1 b_1^* & \dots & a_1 b_n^* \\ \vdots & \ddots & \vdots \\ a_n b_1^* & \dots & a_n b_n^* \end{pmatrix}. \quad (2.6)$$

The density operator of a pure state (for state $|\psi\rangle$), the associated density matrix is $|\psi\rangle\langle\psi|$ could be thought of a function that maps the that pure state to itself. Suppose one of a set quantum pure states $|\phi_1\rangle, \dots, |\phi_n\rangle$ was prepared according to the probability distribution $(p_1, \dots, p_n)$ such that the state was described by the density matrix

$$\rho = \sum_{i=1}^n p_i |\phi_i\rangle\langle\phi_i|. \quad (2.7)$$

If $\{|\phi_i\rangle\}$ are mutually exclusive possibilities, then ρ acts on any particular choice of $|\phi_j\rangle$ to produce $\rho|\phi_j\rangle = p_j|\phi_j\rangle$, and we find that the density matrix has states corresponding to mutually exclusive possibilities as its eigenvectors, and their associated probabilities as its eigenvalues. More remarkably, in quantum theory every state can be written as a probability distribution over a set of mutually exclusive outcomes: a property known as *spectrality* [8]. Equivalently, all density matrices are positive Hermitian matrices with orthogonal eigenvectors and non-negative real eigenvalues.

The density matrix is a more operational description of a system even than writing a probability distribution over kets, in the sense that two states with no observable distinction are described by exactly the same matrix. Consider the example at the top of this section, for instance: an equiprobable distribution over $|0\rangle$ vs $|1\rangle$ would yield *exactly* the same statistics for any conceivable quantum measurement as an equiprobable distribution over $|\psi\rangle$ and $|\psi^\perp\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i(\theta+\pi)}|1\rangle)$. It would even be possible to come up with a larger set of kets (such as a equiprobable distribution over a set of kets with equally spaced $\theta \in [0, 2\pi)$ that yield the same statistics. On the other hand, in all of these cases, the density matrix would be the same: $\mathbb{1}_2/2 = \text{diag}(\frac{1}{2}, \frac{1}{2})$.

Let us consider the form of transformations within the density matrix formulation.

In particular, reversible unitary transformations that would act as $U|\psi\rangle = |\psi'\rangle$ on a pure state written as a ket thus act as $U|\psi\rangle\langle\psi|U^\dagger = |\psi'\rangle\langle\psi'|$ in the density matrix formulation, and more generally act as $\rho \rightarrow \rho' = U\rho U^\dagger$. This can be seen by noting that if the same operation U were applied regardless of the state, a set $\{|\phi_i\rangle\}_i$ with associated probability $\{p_i\}$ would be mapped to the set $\{U|\phi_i\rangle\}_i$ without changing any of the probabilities; and as U preserves inner products, the new set is also a set of mutually exclusive possibilities, and may be written as $\rho' = \sum_i p_i U|\phi_i\rangle\langle\phi_i|U^\dagger = U(\sum_i p_i |\phi_i\rangle\langle\phi_i|)U^\dagger = U\rho U^\dagger$.

Much as classical theory allows for general stochastic matrices as well as permutations, there is a class of irreversible quantum operations. These incorporate the probabilistic application of reversible operations (e.g.) $\rho \rightarrow \sum_j p_j U_j \rho U_j^\dagger$, as well as more general maps (often called ‘quantum channels’) of the form $\rho \rightarrow \rho' = \sum_j B_j \rho B_j^\dagger$ where $\sum_j B_j^\dagger B_j = \mathbb{1}$. This latter restriction imposes a property known as ‘complete positivity’ that guarantees that not only is the output state ρ' a valid density matrix (with non-negative eigenvalues), but also if the map acts on a subsystem in a larger Hilbert space, then the state of this system following the application of the map will also always be a valid density matrix [41, 42]. For this reason, the most general quantum operations⁴ are known as *CP maps*.

2.2.3 Quantum measurements

Let us return to the concept of measurement in quantum theory one more time, to note that the formalism of Hilbert spaces admits a clean way of mapping density matrices to probabilities of measurement outcomes: a process which we shall generalise when it comes to formulating the GPT framework.

First note, in quantum theory, for a set $\{|\phi_i\rangle\}$ to correspond to a complete set of mutually exclusive possibilities, for every quantum state $|\psi\rangle$ applying the Born rule to get the probability associated with each $|\phi_i\rangle$ must sum to unity: $\sum_i |\langle\phi_i|\psi\rangle|^2 = 1$. This means that if $|\psi\rangle = |\phi_j\rangle$, then the sum will contain a term $|\langle\phi_j|\phi_j\rangle|^2 = 1$, and as all the other addends are non-negative, all $|\langle\phi_i|\phi_j\rangle|^2 = 0$, which requires that $\langle\phi_i|\phi_j\rangle = 0$ when $i \neq j$. Hence, a set of mutually exclusive possibilities corresponds to an orthonormal set $\langle\phi_j|\phi_i\rangle = \delta_{ij}$. Such a set obeys a property known as completeness [8] such that $\sum_i |\phi_i\rangle\langle\phi_i| = \mathbb{1}$.

Every quantum state can be written in terms of an arbitrary choice of orthonormal possibilities (and it is for this reason that the choice of possibilities is often referred to as a *basis*), such that a pure state would be $|\psi\rangle = \sum_i c_i |\phi_i\rangle$ for $\{c_i\} \in \mathbb{C}$, and a density

⁴ This also includes maps of the form $\sum_i B_i^\dagger B_i = k\mathbb{1}$ where $0 \leq k \leq 1$, which reduce the normalisation of the state when $k < 1$.

matrix could be expressed $\rho = \sum_{ij} \rho_{ij} |\phi_i\rangle\langle\phi_j|$. From such an expression, it may be seen that the diagonal elements of the density matrix may be selected by $\langle\phi_k|\rho|\phi_k\rangle = \rho_{kk}$ and hence correspond to the probability of outcome $|\phi_k\rangle$.

Hence, for any matrix M expressed in the $\{|\phi_i\rangle\}$ basis, an operation of the form $\sum_i \langle\phi_i|M|\phi_i\rangle$ is equivalent to calculating the *trace* of the matrix $\text{Tr}(M)$. Moreover, the same value will be obtained for any basis choice of mutually exclusive $\{|\phi_i\rangle\}$. Consider an alternative set of outcomes $\{|\psi_i\rangle\}$ which will be related to the first choice by a unitary transformation $|\psi_i\rangle = U|\phi_i\rangle$. For a matrix $M = \sum_{ij} M_{ij} |\phi_i\rangle\langle\phi_j|$, which satisfies $\sum_i \langle\phi_i|M|\phi_i\rangle = \sum_i M_{ii}$, we find

$$\begin{aligned} \sum_k \langle\psi_k|M|\psi_k\rangle &= \sum_k \langle\phi_k|U^\dagger M U|\phi_k\rangle = \sum_k \sum_{ij} M_{ij} \langle\phi_k|U^\dagger|\phi_i\rangle\langle\phi_j|U|\phi_k\rangle \\ &= \sum_{ij} M_{ij} \langle\phi_j|U \left(\sum_k |\phi_k\rangle\langle\phi_k| \right) U^\dagger|\phi_i\rangle = \sum_{ij} M_{ij} \langle\phi_j|U U^\dagger|\phi_i\rangle \\ &= \sum_{ij} M_{ij} \langle\phi_j|\phi_i\rangle = \sum_i M_{ii}, \end{aligned} \tag{2.8}$$

which is the same value. (As a direct corollary, since the unitary operation acting on the set of possibilities could have been expressed as a transformation on the basis of the matrix, we note that the trace of a matrix is invariant under unitary transformations.)

This means, for any outcome $|\phi_j\rangle$ (out of mutually exclusive set of possibilities $\{|\phi_i\rangle\}_i$), the trace $\text{Tr}(|\phi_i\rangle\langle\phi_i|\rho|\phi_i\rangle\langle\phi_i|)$ can be written as $\sum_j \langle\phi_j|\phi_i\rangle\langle\phi_i|\rho|\phi_i\rangle\langle\phi_i|\phi_j\rangle$ which is equivalent to $\langle\phi_i|\rho|\phi_i\rangle$, which is exactly an expression for the probability that a density matrix ρ assigns to the possibility $\langle\phi_i|$ occurring. An expression of the form $P_i = |\phi_i\rangle\langle\phi_i|$ is known as a projector, and in quantum theory is identical to the density matrix associated with the pure state $|\phi_i\rangle$. Combining the Born rule for assigning probabilities to possibilities, along with the update rule following measurements, we note that if a state ρ is measured to be in $|\phi_i\rangle$, then the state following the measurement is exactly the pure state $|\phi_i\rangle\langle\phi_i|$, which can be written in terms of the old state as

$$\rho \rightarrow \frac{P_i \rho P_i^\dagger}{\text{Tr}(P_i \rho P_i^\dagger)} \tag{2.9}$$

(preserving normalisation accounts for the denominator).

This may be naturally generalised beyond projectors, to form the most general type of quantum measurement [8]. For some set of operators $\{M_i\}$, let M_i assign a probability of an outcome labelled i according to $P(M=i) = \text{Tr}(M_i \rho M_i^\dagger) \in [0, 1]$, such that $\sum_i \text{Tr}(M_i \rho M_i^\dagger) = 1$ for all quantum states ρ , and M_i defines an update rule on the

state following the outcome i :

$$\rho \rightarrow \rho_i = \frac{M_i \rho M_i^\dagger}{\text{Tr}(M_i \rho M_i^\dagger)}. \quad (2.10)$$

(We do not need to worry about cases where the denominator is zero, as these correspond to outcomes that never occur.) Such operators necessarily satisfy $\sum_i M_i^\dagger M_i = \mathbb{1}$, which is another form of completeness.

Moreover, we can use the cyclic property of the trace to note that $\text{Tr}(M_i \rho M_i^\dagger) = \text{Tr}(M_i^\dagger M_i \rho)$ and instead write $M_i^\dagger M_i = E_i$, such that the object E_i assigns a probability to ρ according to $\text{Tr}(E_i \rho)$. This formulation is particularly pleasing as it is an inner product, known as the Hilbert-Schmidt inner product. This suggests a natural duality between objects like E_i and between density matrices ρ (much as bras and kets had a duality defined by an inner product). Choosing a set of $\{E_i\}$ defines a very general class of quantum measurement, known as *POVMs* (positive operator-valued measures). It can be shown (e.g. §2.2.6 in [8]) that for *any* set of positive matrices (i.e. matrices without negative eigenvalues) $\{E_i\}$, which satisfy $\sum_i E_i = \mathbb{1}$, a set of measurement operators M_i can always be found such that $E_i = M_i^\dagger M_i$, and if we are predominantly concerned with the assignment of probabilities, rather than precise post-measurement states, then $\{E_i\}$ is a sufficient specification for the measurement.

2.3 The framework of generalised probabilistic theories (GPTs)

So far we have briefly reviewed the classical description of the states of a system using probability vectors, and contrasted this with the quantum description which uses complex Hilbert spaces. In both of these descriptions, there were shared concepts such as states, transformations and measurement; although the details of how these concepts worked in each framework proved somewhat different. The *generalised probabilistic theory* (GPT) framework, which we shall introduce throughout the remainder of this chapter (see also [9, 10, 14, 18, 43–46]), attempts to put these theories on an even footing by abstracting these properties in such a way that they may be expressed in a shared mathematical structure.

2.3.1 The generalised experiment

We can break the operational description of an experiment into three sections [9, 14] (broadly reflecting the divisions outlined in chapter 2 of [47]). We shall describe these stages in general, and for didactic purposes also invoke the example of a quantum Mach-Zehnder interferometer drawn in fig. 2.4.

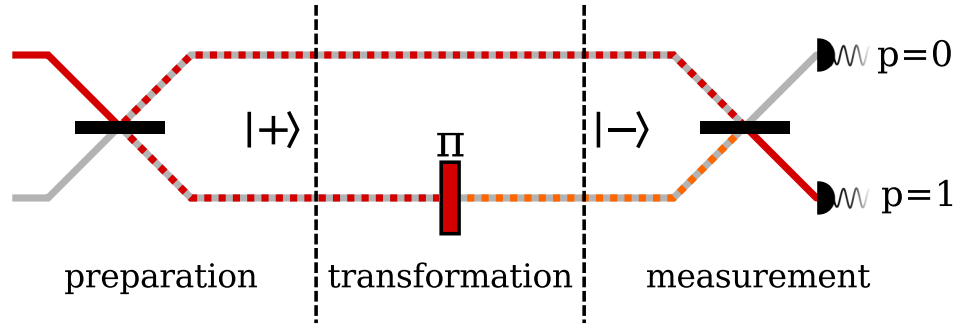


Figure 2.4: A Mach-Zehnder interferometer In this set-up, a single photon may occupy one of two branches. The photon is incident on the first beam-splitter, and enters into a superposition of the two branches. A phase-plate (a component that changes the path length, such as an extra coil of optical fibre, or piece of glass) is introduced on one of the branches, which changes the phase between the branches. Finally, the two branches recombine with a second beam-splitter, and the single photon arrives probabilistically in one of the two detectors. In the figure drawn here, the phase induced is π , which changes a quantum superposition $|+\rangle = |0\rangle + |1\rangle$ into $|-\rangle = |0\rangle - |1\rangle$, such that the action of the second beamsplitter is to send the photon into the lower detector with certainty.

1. **Preparation.**— In the preparation stage, the experimentalist tunes some dials, then pushes a button to release an object, which if measured will manifest outcomes according to a set of statistics specified by the dial settings at the time of creation. Consider in the MZI: a photon initially produced in upper branch is incident on a beam-splitter, such that afterwards, it is in a particular superposition of upper and lower branches- this entire composite process may be considered the *preparation* of the state $|+\rangle$.
2. **Transformation.**— In the transformation stage, the experimentalist acts upon the system in a way such that the current state of the system is updated to take this action into account. In the MZI example, this is achieved by adding a phase plate (piece of glass) in one of the arms, to alter the relative phase between the branches; but in general this could be any action, including inserting another beam-splitter, or even swapping the paths around.
3. **Measurement.**— In the measurement stage, the experimentalist chooses a way in which to probe the system, which results in the output of classical statistics

(i.e. the relative frequency of possibilities). In the MZI interferometer the final beamsplitter can be seen as part of the composite process to measure in the basis $|+\rangle$ vs. $|-\rangle$ (see also fig. 2.3 in which different choices of beamsplitter effect different measurements on the system); here, the relative frequency of clicks in either detector may be described by a classical probability vector, and hence constitutes an output of classical statistics.

2.3.2 States

Performing a measurement may be thought of as choosing a set of possibilities in a system (e.g. $\{\text{“detector one clicks”}, \text{“detector two clicks”}\}$), and then resolving the likelihood of obtaining each possibility into a classical probability distribution. Suppose we have a set of possibilities $\{x_1, \dots, x_n\}$ and label this choice of set as measurement X . Then, we would write the probability that possibility x_i (also known as an **outcome** in the context of a measurement) occurs as $P(X = x_i)$ or $P(x_i|X)$.

For a particular system, it could be that choosing different measurements results in very different output probability distributions. However, within the GPT framework, we shall consider theories where it is possible to find enough different measurements, such that by specifying the statistics of all these measurements, it is possible to determine the statistics of *any* conceivable measurement on the system. Indeed, in an operational approach, there is nothing else to know about the system other than these statistics; and so this collection of statistics itself forms the mathematical representation of the state of the system.

Definition 2.1: State

The **state** of a system is given by the complete set of statistics necessary to completely determine the classical probability distribution for any measurement that could be made on the system.

Recall that for a classical system (which, excepting coarse-graining, only admits one choice of measurement), the states are represented by probability distributions, and in the quantum case states could be represented by *density matrices*. In the GPT formalism, the state can be represented as a collection of probability distributions for the minimum number of independent measurements required to encompass everything known about the system. The outcome probabilities for each measurement can be written as a vector of real numbers, which can be combined naïvely to form a larger

vector of real numbers [18, 48]:

$$\mathbf{s} = \begin{bmatrix} P(X = x_1) \\ \vdots \\ P(X = x_n) \\ \hline \vdots \\ P(Z = z_1) \\ \vdots \\ P(Z = z_m) \end{bmatrix}, \quad (2.11)$$

where we have used the upper case symbols X, Z to represent (two of) the different measurements, which in general could have a different number of outcomes.

A state is said to be **normalised** if any measurement performed on it is guaranteed to give an outcome (i.e. all the classical probability distributions assigned to the different measurements are themselves normalised). This places the restriction on the normalisation constant $\mathcal{N}$,

$$\mathcal{N} = \sum_i P(X = x_i) = 1 \quad (2.12)$$

for every measurement X . If this restriction does not hold, the state is *not normalised*, and there is a chance that the state ‘disappears’ when a measurement is made on it. Within the GPT framework, we only admit states where the value of the normalisation constant $\mathcal{N}$ is independent of the choice of measurement, such that the choice of measurement does not affect how likely a system is to exist. (Why this should be the case is not obvious when just considering a single system, but can be motivated by a physical restriction discussed in § 2.6).

2.3.3 State-spaces

State-spaces in the GPT framework are defined much like they are in classical theory:

Definition 2.2: State-space

The **state-space** of a system is defined to be the set of all allowed states of that system.

Certainly, every state must be comprised of valid probability distributions, but in some cases this is not the only restriction on states. For example, in quantum theory, every pure state is associated with a unique projective measurement outcome that returns success with probability 1. This means there is only one state where (say) $P(Z = z_1) = 1$, and hence there is no freedom to choose how $P(X = x_1), P(X = x_2)$,

etc. could be assigned.

However, there is one rule that every system describable by the GPT framework must satisfy: the probabilistic mixtures of allowed states are also allowed states. For a preparation of state s_1 with probability $\lambda \in [0, 1]$ or s_2 with probability $(1 - \lambda)$, by applying standard probability theory $[P(B) = \sum_i P(A_i)P(B|A_i)]$ to each element of the vector, the mixed state s' is given by the convex combination⁵:

$$s' = \lambda s_1 + (1 - \lambda) s_2. \quad (2.13)$$

This implies by definition that the state-space is a convex set. It is for this reason that the GPT formalism is sometimes known as the **convex framework**.

We can use this as a natural way to define the mutually-exclusive categories of pure and mixed states within the GPT framework:

Definition 2.3: Pure state

A state is **pure** if it cannot be formed by the probabilistic combination of other states in the state-space.

Definition 2.4: Mixed state

A state is **mixed** when it can be formed by the probabilistic combination of other states in the state-space.

Finally we remark that when we are only concerned with binary measurements on a normalised state⁶, it is possible to use an **expectation value representation** to express the statistics of the state. If we label the two outcomes of each measurement as $+1$ and -1 , we can write

$$\langle X \rangle = P(X = +1) - P(X = -1).$$

and then frame the expectation values for the different choices of measurement as a vector. Due to their binary nature, this allows us to form a meaningful geometric representation in Euclidean space (of dimension equal to the number of measurements used to specify the state), in which for each measurement, one outcome corresponds to a different Cartesian axis, and each point within corresponds to a state with different statistics [49]. Most of the diagrammatic representations of state-spaces in this thesis will follow this procedure.

⁵ It is possible to formulate a more abstract definition of convexity (see §2.2 in [45]), but the function we adopt here is particularly useful for probabilities.

⁶ Moreover, for any measurement X with a finite number of outcomes N , one can find $\lceil \log_2 N \rceil$ binary measurements that cover all the statistics of X .

2.3.4 Effects

Let us now formalise the concept of ‘possibilities’ with a mathematical object known as an *effect*. Much like how in quantum theory a bra maps a ket onto a complex number, an effect is a functional that maps a state into a single probability. A set of possibilities in a measurement are thus represented by a set of effects, which will map that state onto a classical probability distribution.

Definition 2.5: Effect

Associated with a every measurement outcome is an **effect**: a functional which acts on a state to assign the *probability* that that outcome would occur, should that measurement be made.

When a states is expressed as a real vector, an effect associated with outcome x_i of measurement X may be expressed as a covector written $\mathbf{x}_i$ such that the associated measurement outcome probability for state $\mathbf{s}$ is given by the inner product, $\mathbf{x}_i \cdot \mathbf{s} = P(X = x_i)$. (This covector is also real [48]). For example, for a state expressed in the form of eq. (2.11), the effect associated with outcome x_1 of measurement X would be written:

$$\mathbf{x}_1 = [1, 0, \dots, 0 \mid \dots \mid 0, \dots, 0]. \quad (2.14)$$

In quantum theory, the set of states and effects are naturally duals of each other (such that for every pure state $|\psi\rangle$, there is a single bra $\langle\psi|$, associated with the outcome of a fine-grained measurement which occurs with certainty on the state $|\psi\rangle$), this does not have to be the case in general. Hence, in the GPT framework we are more cautious about treating states and effects as conceptually separate objects than we would need to be in quantum or classical theory.

We remark that it is possible to **coarse-grain** a measurement X with n outcomes (and hence n effects), $\{\mathbf{x}_1 \dots \mathbf{x}_n\}$ into a measurement X' with $n' < n$ effects $\{\mathbf{x}'_1 \dots \mathbf{x}'_{n'}\}$ by generating a new set of effects by summing together combinations of elements in the original set, such that every effect is included once and only once. For example,

$$\begin{aligned} \mathbf{x}'_1 &= \mathbf{x}_1 + \dots + \mathbf{x}_{k_1}, \\ \mathbf{x}'_2 &= \mathbf{x}_{k_1+1} + \dots + \mathbf{x}_{k_2}, \\ &\vdots \\ \mathbf{x}'_{n'} &= \mathbf{x}_{k_{n'-1}+1} + \dots + \mathbf{x}_n. \end{aligned} \quad (2.15)$$

In this case, if $\mathbf{x}_1 \cdot \mathbf{s} = p_1$ and $\mathbf{x}_2 \cdot \mathbf{s} = p_2$, then $(\mathbf{x}_1 + \mathbf{x}_2) \cdot \mathbf{s} = p_1 + p_2$.

By taking coarse-graining to its ultimate extreme, one could consider a measurement with a single effect outcome, such that when this effect is applied to the state, it

returns the normalisation of the state. This effect is known as the **unit effect** and is written $\mathbf{1}$. There is no unique representation of this, but because the normalisation of states in the GPT framework are guaranteed to be the same for every measurement it is safe to write as [18], say

$$\mathbf{1} = (1, \dots, 1 \mid 0, \dots, 0 \mid \dots \mid 0, \dots, 0) \quad (2.16)$$

such that $\mathbf{1} \cdot \mathbf{s} = \sum_i P(X = x_i) = \mathcal{N}$ for all $\mathbf{s}$.

We can also use the formalism of effects to generate *expectation values*. By assigning a real-valued number x_i to each measurement outcome labelled by effect $\mathbf{x}_i$, then for a state $\mathbf{s}$:

$$\langle X \rangle = \sum_i x_i P(X = x_i) = \sum_i x_i \mathbf{x}_i \cdot \mathbf{s} \quad (2.17)$$

and so the object $\mathbf{X} = \sum_i x_i \mathbf{x}_i$ is a functional which maps states onto expectation values. This is particularly useful for binary measurements, where (choosing as the labels $+1$ and -1), we find the functional $\mathbf{E} = \mathbf{e}_+ - \mathbf{e}_-$. As $\mathbf{e}_+ + \mathbf{e}_- = \mathbf{1}$, knowing $\mathbf{E} \cdot \mathbf{s}$ alongside the normalisation $\mathcal{N}$ is sufficient to completely determine the statistics for that measurement.

2.3.5 Transformations

Definition 2.6: Transformation

A **transformation** is a map between states of a system.

As the state of the system is specified in terms of the statistics, transformations are maps between sets of statistics.

A transformation is **valid** on a state-space if it maps every state in that state-space to an allowed state (such that there are no output states not contained within the state-space allowed for that system). More strongly, a transformation is **allowed** if it is valid, and does not violate any other restriction that may have been explicitly placed.

Within the GPT framework transformations are taken to be linear, which we motivate by a property known as *linearity of mixing*. Consider a transformation $f : \mathbf{s} \rightarrow f(\mathbf{s})$. If the particle is in one of two states $\mathbf{s}_1$ or $\mathbf{s}_2$ with probabilities λ and $(1 - \lambda)$ before the transformation, then this transformation will act on each state $\mathbf{s}_1 \rightarrow f(\mathbf{s}_1)$ and $\mathbf{s}_2 \rightarrow f(\mathbf{s}_2)$. Linearity of mixing means that if we were to apply the transformation to the combined state, $\mathbf{s} = \lambda \mathbf{s}_1 + (1 - \lambda) \mathbf{s}_2$ such that $\mathbf{s} \rightarrow f(\mathbf{s})$, then $f(\mathbf{s})$ should be the same as the result we get by taking the probabilistic mixture of the outputs $\lambda f(\mathbf{s}_1) + (1 - \lambda) f(\mathbf{s}_2)$, implying:

$$f(\lambda \mathbf{s}_1 + (1 - \lambda) \mathbf{s}_2) = \lambda f(\mathbf{s}_1) + (1 - \lambda) f(\mathbf{s}_2) \quad (2.18)$$

for all $\lambda \in [0, 1]$. This restriction implies that f is an *affine transformation*, meaning we can represent transformations as matrices pre-multiplying the vectors of probabilities.

Of particular interest are the class of transformations which are reversible.

Definition 2.7: Reversible transformation

An allowed transformation $T : s \rightarrow s'$ is called **reversible** if there is also a (possibly different) allowed transformation T^{-1} which maps every output state of T back onto the corresponding input state that generated it, and moreover T^{-1} is an allowed transformation for *all* states in the state-space.

It follows that the image of a state-space through a reversible transformation is the same state-space, and so reversible transformations form a group under composition. This group will be the *automorphism group* of the state-space, or some subgroup thereof if there are other restrictions on which transformations are admitted.

Transformations not satisfying the criterion of reversibility are known as *irreversible transformations*. We remark that although a transformation described by a singular (non-invertible) matrix will be irreversible, singularity is not necessary for irreversibility. For instance, a uniform shrinking of the state-space may be performed in a non-singular way by a matrix T (for example, mapping every probability in the range $[0, 1]$ to the range $[\frac{1}{4}, \frac{3}{4}]$ via the map $p' = \frac{1}{4} + \frac{1}{2}p$) but if the mathematical inverse T^{-1} of this matrix were applied (mapping $p = 2(p' - \frac{1}{4})$), states outside of the image of the state-space through T would be mapped by T^{-1} to states outside the state-space (e.g. a state with an element where $p' = 0$ would be mapped to $p = -0.5$ which is not a valid probability).

Finally, we define what it means for a transformation to be trivial.

Definition 2.8: Trivial transformation

A transformation is *trivial* if it maps every state within a state-space to itself.

By its very definition, the trivial transformation corresponds to an *identity* operator, and within the vector representation, this can be represented by a identity matrix (explicitly a diagonal matrix with only 1s along the diagonal, written here as $\mathbb{I}_n$ where n is the dimensionality of the matrix). However, when states are represented as in eq. (2.11), they are over-determined because of the constraint that all measurements normalise in the same way. Thus, much like the unit effect 1 , one could construct other matrices than the identity that have no effect on any allowed state. This definition of trivial is much stricter than that used by Gross et al. [50], which incorporates a wide range of reversible transformations corresponding to the relabelling of measurements, measurement outcomes, or subsystems (discussed in § 2.6), but excludes transformations that generate correlations.

2.3.6 Theories

Finally, we bring the above components together to form a theory.

Definition 2.9: Theory

A **theory** assigns to a physical system (set-up), a complete description of its operational behaviour. This description consists of a tuple of sets of:

- i. all allowed *effects* (every possible measurement and outcome),
- ii. all allowed *states* (every possible combination of statistics),
- iii. all allowed *transformations* (every allowed way in which the statistics can change).

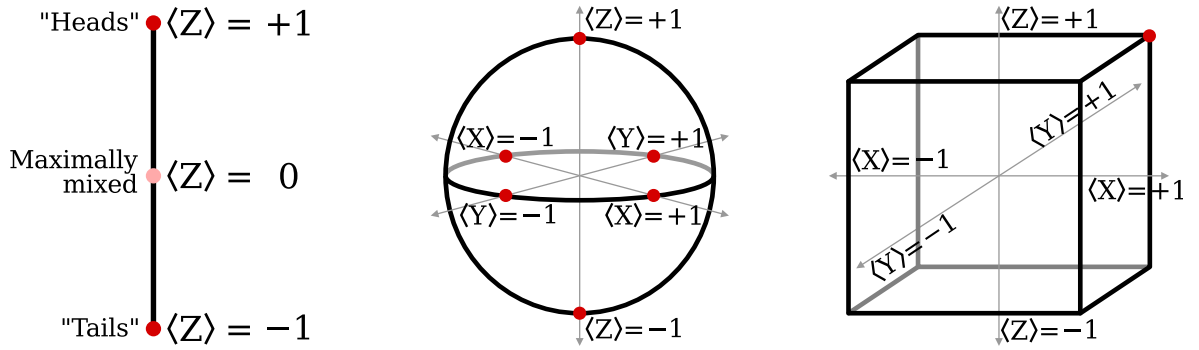
For any given state-space, it is possible to suggest a set of valid effects (the dual space of functionals that would map the states onto the range $[0, 1]$), and a set of transformations (by considering every automorphism of the state-space, and the convex combination thereof). However, the state-space alone is not sufficient to completely define the theory, as there could be further arbitrary restrictions placed on the allowed effects and transformations restricting the choice to a subset of the effects and transformations found in the above manner. Thus, a set-up in two different theories could share the same state-space, but admit alternative transformations.

2.4 Example theories within the GPT framework

Let us now return to a few concrete examples, to see how they fit in the GPT framework. We will revisit classical and quantum theory, before moving onto a theory that goes beyond quantum theory. So that these examples are more directly comparable, in all cases we will consider binary systems, in which the theory describes the state of a system with two mutually-exclusive possibilities.

2.4.1 Classical probability theory: *bits*

Classical theories only admit one choice of measurement, and so the states as they might be expressed within the GPT framework (definition 2.1) are exactly the same as the probability vectors as described in §2.1. The effects associated with the possibility i may be written as covectors $\mathbf{i} = (0, \dots, 0, 1, 0, \dots, 0)$ such that all elements are 0 except column i . Transformations are exactly the set of stochastic matrices, which map probability distributions to probability distributions.



(a) Classical bit. The two red points at the ends correspond to pure states, where the state of the coin is known to be definitely heads, or definitely tails. The mid-point of the line represents the maximally mixed state, such as what an observer might assign if a fair coin was tossed and the result withheld.

(b) Qubit. The state-space of a single qubit is the Bloch sphere. Every point on the surface corresponds to a pure state of a qubit, and every point inside corresponds to a mixed state. There are three complementary choices of measurement basis. The points in red correspond to pure states associated with the eigenvectors of the Pauli matrices.

(c) 3-in 2-out gbit– (definition 2.10) all three measurements can independently take maximal values; and so there is a plane of states for each definite measurement outcome (e.g. $\langle Z \rangle = 1$), and the state-space forms a cube. This state-space admits non-quantum states, such as the point in red where the outcome of every measurement is +1.

Figure 2.5: Three example state-spaces.

The drawn example (fig. 2.5a) shows a classical bit: a binary choice. If we label the only measurement as Z , and associate ‘heads’ with $+1$ and ‘tails’ with -1 , we would write the general state of this system as:

$$\mathbf{s} = \begin{pmatrix} P(Z = +1) \\ P(Z = -1) \end{pmatrix}, \quad (2.19)$$

and the effects $\mathbf{z}_+$ and $\mathbf{z}_-$ associated with heads and tails respectively would be written:

$$\mathbf{z}_+ = (1, 0), \quad \mathbf{z}_- = (0, 1). \quad (2.20)$$

The allowed transformations on the states are given by 2×2 stochastic matrices of the form

$$T = \begin{pmatrix} p & 1 - q \\ 1 - p & q \end{pmatrix}, \quad p, q \in [0, 1]. \quad (2.21)$$

The two reversible transformations correspond to the choices $p = q = 1$ and $p = q = 0$. The former corresponds to the trivial transformation $\mathbb{1}$, and the latter to exchanging

$+1$ and -1 : a bit flip. Under composition, these matrices form a group isomorphic to the integer modulo-2 addition group Z_2 .

2.4.2 Quantum theory: *qubits*

Recall from § 2.2.3 that there was an inner product function $\text{Tr}(E_i \rho)$ that assigned probabilities to some outcome labelled by E_i to a state labelled by ρ . This is naturally isomorphic to the inner product structure between effects and states in the GPT framework where $\mathbf{e} \cdot \mathbf{s}$ assigns a probability of an outcome labelled $\mathbf{e}$ to a state labelled by $\mathbf{s}$. The CP maps are the transformations.

Let us consider the specific example of a quantum binary system: the qubit. Recall from § 2.2 that the general state of a two-level quantum system is given by a 2×2 density matrix, which will satisfy:

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} = \begin{pmatrix} a & b - ic \\ b + ic & 1 - a \end{pmatrix}, \quad (2.22)$$

where the expression on the right-hand explicitly shows the three real degrees of freedom remaining after we note that $\rho = \rho^\dagger$ and that $\text{Tr}(\rho) = 1$.

Let us now consider a few quantum measurements that would be sufficient to determine the state. First, we take the ‘computational basis’ measurement, labelled Z , with outcomes $+1$ and -1 associated with $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$ respectively. For these, $P(Z = +1) = \text{Tr}(|0\rangle\langle 0| \rho) = a$ and $P(Z = -1) = \text{Tr}(|1\rangle\langle 1| \rho) = 1 - a$.

Similarly, we can choose two other measurements X and Y such that:

$$\begin{aligned} X^+ &= |+\rangle\langle +| = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} & X^- &= |-\rangle\langle -| = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \\ Y^+ &= |i\rangle\langle i| = \frac{1}{2} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} & Y^- &= |-i\rangle\langle -i| = \frac{1}{2} \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \end{aligned} \quad (2.23)$$

such that expressing the probabilities for all three measurements gives

$$\mathbf{s} = \begin{pmatrix} P(X = +1) \\ P(X = -1) \\ \frac{P(Y = +1)}{P(Y = -1)} \\ \frac{P(Z = +1)}{P(Z = -1)} \end{pmatrix} = \begin{pmatrix} 1/2 + b \\ 1/2 - b \\ \frac{1/2 + c}{1/2 - c} \\ a \\ 1 - a \end{pmatrix}. \quad (2.24)$$

Clearly $\mathbf{s}$ encodes all of a , b and c and so is sufficient to specify the qubit. Thus $\mathbf{s}$ qualifies as a state in the GPT framework. The effects associated with the X , Y and Z

measurement outcomes have obvious representations, given by

$$\begin{aligned} \mathbf{x}_+ &= (1, 0 | 0, 0 | 0, 0), & \mathbf{x}_- &= (0, 1 | 0, 0 | 0, 0), \\ \mathbf{y}_+ &= (0, 0 | 1, 0 | 0, 0), & \mathbf{y}_- &= (0, 0 | 0, 1 | 0, 0), \\ \mathbf{z}_+ &= (0, 0 | 0, 0 | 1, 0), & \mathbf{z}_- &= (0, 0 | 0, 0 | 0, 1). \end{aligned} \quad (2.25)$$

We remark that if we combine the effects to form new functionals, $\mathbf{X} = (\mathbf{x}_+ - \mathbf{x}_-)$, $\mathbf{Y} = (\mathbf{y}_+ - \mathbf{y}_-)$ and $\mathbf{Z} = (\mathbf{z}_+ - \mathbf{z}_-)$, from the linearity of the inner product we see that these return the expectation value of the respective measurements weighting numerically between the labels $+1$ and -1 : $\mathbf{Z} \cdot \mathbf{s} = 2a - 1$, $\mathbf{X} \cdot \mathbf{s} = b$ and $\mathbf{Y} \cdot \mathbf{s} = c$. This is exactly analogous to taking the quantum expectation value⁷ with the three Pauli matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (2.26)$$

which is the motivation behind our choice of labels X , Y and Z . As the state $\mathbf{s}$ can be determined completely by these three expectation values, we may draw the (normalised) state-space of a qubit in three dimensions as in fig. 2.5b (see also the end of §2.3.3).

The spherical shape of the qubit state-space arises from a restriction on which statistics correspond to valid quantum states known as the (quantum) *uncertainty principle*. This uncertainty principle follows as a direct consequence of the structure of Hilbert spaces (see appendix A). For qubits, this restriction is written

$$\langle \sigma_x \rangle^2 + \langle \sigma_y \rangle^2 + \langle \sigma_z \rangle^2 \leq 1. \quad (2.27)$$

It may be seen that whenever the state has a certain outcome for one of the X , Y or Z measurements, the outcomes for the other two measurements are totally random. This property is deducible entirely from the restrictions on the state-space, and thus operationally we will treat the quantum uncertainty principle as a property of the state-space.

As a sphere, the automorphisms of the qubit state-space are given by the three-dimensional orthogonal group $O(3)$. Quantum theory does not admit any operation that reflects only one of the axes (e.g. $\langle Y \rangle \rightarrow -\langle Y \rangle$), as this would correspond to a disallowed map $\rho \rightarrow \rho^*$, which is not completely positive. Thus, the set of allowed reversible transformations on a qubit are given by the subgroup of proper rotations:

⁷ Converting quantum measurement operators into an object that generates the expectation value works just as it does for GPT effects. Noting that the trace is a linear operation, assigning a value m_i to a measurement outcome E_i for a measurement E , the quantum expectation value is given by $\sum m_i P(m_i | E) = \sum m_i \text{Tr}(E_i \rho) = \text{Tr}((\sum_i m_i E_i) \rho)$ where we can write $E = \sum_i m_i E_i$, such that $\langle E \rangle = \text{Tr}(E \rho)$. If all labels m_i are real numbers, then E will be Hermitian.

the special orthogonal group $SO(3)$. There is a homomorphism from $U \in U(2)$ which acts on the qubit Hilbert space, to $T \in SO(3)$ acting on the real-valued GPT framework state. This morphism has an additional $U(1)$ freedom in that U and $e^{i\phi}U$ map to the same T , but because ρ transforms as $\rho \rightarrow U\rho U^\dagger$, this extra freedom is not observable on an isolated qubit, and hence not of operational importance for describing the allowed qubit behaviour.

2.4.3 Beyond quantum theory: *gbits*

Now let us step beyond quantum theory, and consider a theory which admits the possibility for alternative choices of measurement, but is not backed by the well-established mathematical structure of Hilbert spaces.

A simple choice of such a theory is given by **gbits**. Gbits are a set of theories in which there are no restrictions on measurement outcomes, except those intrinsic to the GPT framework: that the measurements all normalise to the same value, and that probabilities are bound between 0 and 1. For state-spaces where every measurement used to specify a state has the same number of outcomes, we will borrow notation from the device-independent cryptography community, and refer to *m*-in *n*-out gbits.

Definition 2.10: m-in n-out gbits

An **m-in n-out gbit** is the state-space described by *m* different measurements, each with *n* mutually exclusive outcomes, with no restrictions on the statistics each measurement can take.

The “*m*-in *n*-out” notation comes from a context in which one imagines an agent with free will in choice of *m* possible inputs to a black-box device. The black-box then returns one of *n* outcomes (with labels comparable between different choices of *m*), which depend on the choice of input *n* and the current state of the system. In the GPT framework, we associate choosing a measurement with choosing the input of such a device. Gbits are the single system state-spaces in a larger framework known as **box-world**. (The multi-partite systems in box-world will be discussed further in §2.6).

Of particular interest is the 3-in 2-out gbit, as drawn in fig. 2.5c. Such a theory forms a natural foil to the qubit, as it is also described by three binary measurements, but in the gbit case these measurements are not subject to an uncertainty restriction.

This allows the states such as

$$\mathbf{s} = \begin{pmatrix} \frac{P(X = +1)}{P(Y = +1)} \\ \frac{P(X = -1)}{P(Y = -1)} \\ \frac{P(Z = +1)}{P(Z = -1)} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad (2.28)$$

which are not possible in a quantum state-space. The effects for X , Y and Z measurements as described for the qubit in § 2.4.2 are valid on gbits, as are the expectation values of these measurements. However, not every valid quantum effect is also a valid gbit effect. For instance the effects associated for a qubit measurement along an axis half way between X and Z would be invalid for gbits, because these would assign probabilities greater than 1 or less than 0 to some gbit states (such as $\mathbf{s}$ in eq. (2.28)).

An interesting pair of effects for a cubic gbit is given by:

$$\mathbf{e}_+ = \frac{1}{3}(1, 0 | 1, 0 | 1, 0), \quad \mathbf{e}_- = \frac{1}{3}(0, 1 | 0, 1 | 0, 1), \quad (2.29)$$

which could be interpreted as a measure across the corners of the cube, or alternatively measuring one of X , Y or Z with equal probability, recording the outcome, but ‘forgetting’ which of X , Y or Z was chosen. These effects constitute a measurement (even in quantum theory), as $\mathbf{e}_+ \cdot \mathbf{s} + \mathbf{e}_- \cdot \mathbf{s} = \mathcal{N}$ for all $\mathbf{s}$. For the gbit state $\mathbf{s}$ in eq. (2.28), $\mathbf{e}_+ \cdot \mathbf{s} = 1$ and $\mathbf{e}_- \cdot \mathbf{s} = 0$. However, there are no quantum states $\mathbf{q}$ where $\mathbf{e}_+ \cdot \mathbf{q} = 1$, and so $\{\mathbf{e}_+, \mathbf{e}_-\}$ cannot be a projective quantum measurement.

The name gbit comes historically from ‘generalised-bit’ (in that the gbit admits more choice of states than a qubit); but in the context we described here this name is misleading, as the gbit is only one specific example of many in the GPT framework. (Also, an “ m -in n -out gbit” for $n > 2$ is technically a ‘dit’ rather than a bit, a word which is itself a contraction of binary digit.) In some literature, 3-in 2-out gbits may be referred to as a cubic bit (or the phonetically ambiguous contraction ‘cubit’), and 2-in 2-out gbits as a square bit (or ‘squit’). These names are in reference to the shape of the state-space they define. In general, an m -in 2-out gbit has a m -dimensional hypercube as its state-space.

It should be noted that whilst an m -in n -out gbit allows for more choices of state than an n -level system in quantum theory would, it still only permits at most n mutually exclusive outcomes for any choice of measurement (this distinction between the *degrees of freedom* needed to specify a state, and the *dimension* corresponding to mutually exclusive possibilities forms the basis of Hardy’s axiomatic derivation of quantum theory [9]).

Typically, we shall take the reversible transformations on a gbit to be exactly the automorphism group of its state-space. Unlike quantum theory, this will be a finite-group. Operations in this group include permutations of the labelling of measurements, permutations of the labelling of outcomes, and combinations of the two. For m -in 2-out gbts, this group will be the symmetries of an m -dimensional cube. Occasionally for hypercubic gbts, we might consider a variant theory in which we only admit reversible transformations corresponding to the proper rotation subgroup of automorphisms. This would be done in order to make the gbit more similar to quantum theory, which also disallows reflections.

2.5 Minimum-dimension representation

As discussed in §2.3.2, the GPT framework requires all measurements to normalise to the same value. Consider the most general state, specified in terms of N measurements $X_1 \dots X_N$, each with a (possibly different) number of respective outcomes $x_1^1 \dots x_{M_1}^1$ to $x_1^N \dots x_{M_N}^N$. Once the choice of $\mathcal{N} = \sum_i P(X_1 = x_i)$ is fixed, the other measurements are constrained to satisfy $\mathcal{N} = \sum_i P(X_2 = x_i) = \dots = \sum_i P(X_N = x_i)$. Thus if the normalisation $\mathcal{N}$ of a state is known, it is not necessary to specify the probability of the final outcome of each measurement. We can therefore write the state as a vector of the form (introduced in [1]):

$$\mathbf{s} = \begin{pmatrix} \frac{\mathcal{N}}{P(X_1 = x_1^1)} \\ \vdots \\ \frac{P(X_1 = x_{M_1-1}^1)}{P(X_1 = x_1^1)} \\ \vdots \\ \frac{P(X_N = x_1^N)}{P(X_N = x_1^N)} \\ \vdots \\ \frac{P(X_N = x_{M_N-1}^N)}{P(X_N = x_1^N)} \end{pmatrix}. \quad (2.30)$$

A vector of this form is useful because in the absence of further constraints, it contains as many dimensions as real degrees of freedom in choice of state, $1 + \sum_{i=1}^N (M_i - 1)$. In such a representation, the unit effect is expressed uniquely as:

$$\mathbf{1} = (1 | 0, \dots, 0 | \dots | 0, \dots, 0). \quad (2.31)$$

Likewise, the only matrix which (in the absence of additional constraints on the state-space) does not change anything is the identity matrix $\mathbb{1}$, with ones in the diagonal and zeroes elsewhere.

As a simple example, consider a state-space describable by two binary measurements $X = \pm 1$ and $Z = \pm 1$. The minimum-dimension representation of this is:

$$\mathbf{s} = \left(\frac{\mathcal{N}}{\frac{P(X = +1)}{P(Z = +1)}} \right). \quad (2.32)$$

By knowing such a vector, we can calculate the remaining probabilities:

$$\begin{aligned} P(X = -1) &= \mathcal{N} - P(X = +1), \\ P(Z = -1) &= \mathcal{N} - P(Z = +1), \end{aligned} \quad (2.33)$$

which would be associated with the respective effects

$$\begin{aligned} \mathbf{x}_- &= \mathbf{1} - \mathbf{x}_+ = (1 \mid -1 \mid 0), \\ \mathbf{z}_- &= \mathbf{1} - \mathbf{z}_+ = (1 \mid 0 \mid -1). \end{aligned} \quad (2.34)$$

2.6 Multi-partite systems in the GPT framework

When considering composite or *multi-partite* (composed of smaller parts) systems, there are two approaches to the problem. The first approach is from the ground up: one considers multiple copies of single systems, and then lists the possible ways in which the statistics of the single systems may be correlated with each other. The second is from the top down: one considers a single system that corresponds to the composite system in its entirety, and then identifies which parts belong to various subsystems. The other extreme choice of multi-partite state-spaces then is the set of all states which return valid probabilities when a joint (i.e. simultaneous) measurement is made on the composite single systems.

When starting from the ground up, one will encounter a highly non-trivial problem, corresponding to finding the ‘tensor product’ between convex state-spaces: the set of allowed correlations between the two systems. Let us consider the two extremes (known as the min and max tensor products [44], a well-studied concept in monoidal category theory, and C^* -algebras [51]).

The smallest allowed state-space assumes that both of the subsystems can independently take any allowed single-system state, and so the joint state-space consists of the convex closure of these ‘product states’ over a classical probability distribution. For example, in quantum theory such a procedure would generate the set of *separable quantum states*.

On the other hand, there could be more extreme correlations than possible by this construction. For example, in quantum theory *entangled states*, such as the state

$|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ on two qubits, cannot be expressed as a classical probability distribution over independently prepared single systems [52, 53].

The other extreme choice of multi-partite state-spaces then is the set of all states which return valid probabilities when a joint (i.e. simultaneous) measurement is made on the composite single systems. (This is within the set of all vectors that can be ‘projected’ onto a valid single-system state-space.) In the quantum example, $|\Psi\rangle$ satisfies this definition. Any measurement made on just one of the parts will be as if the maximally mixed state $\rho = \frac{1}{2}\mathbb{1}_2$ were measured, which is a valid quantum state. (Moreover, these states must not violate a property known as *no-signalling*, which we will discuss shortly.)

If one takes the top-down approach, it amounts to assuming that the valid state-space has already been decided upon, and one has to identify operationally what about that state-space constitutes being in one subsystem over another. This can be done by choosing non-overlapping subsets of the allowed effects that correspond to the set of measurements and outcomes that are observable by an agent in each individual subsystem.

We are particularly interested in set-ups where one set of effects accounts for what is measurable by an agent located at some point in space, and another set corresponds to what might be measured by a different agent at a separate location. In such a set-up, theories in the GPT framework follow a physical restriction known as *no-signalling*: that is, they do not permit communication faster than the speed of light. This restriction may be formalised as follows [54]: Suppose there are two spatially disjoint agents Alice and Bob, and they both have access to a bipartite state. No choice of measurement (A vs. A') made by Alice should affect the outcomes of any possible measurement B by Bob, and vice versa:

$$\begin{aligned} \sum_i P(\mathbf{a}_i, \mathbf{b}_j | A, B) &= \sum_i P(\mathbf{a}_i, \mathbf{b}_j | A', B) \quad \forall \mathbf{b}_j, A, A', B, \\ \sum_j P(\mathbf{a}_i, \mathbf{b}_j | A, B) &= \sum_j P(\mathbf{a}_i, \mathbf{b}_j | A, B') \quad \forall \mathbf{a}_i, A, B, B', \end{aligned} \quad (2.35)$$

where $P(a_i, b_j | A, B)$ is the joint probability of Alice getting outcome a_i when she makes measurement A , and Bob getting outcome b_j when he makes measurement B . (A corollary of this restriction is that all measurements must share the same normalisation, as asserted in §2.3.2).

We can define the **reduced states** of a system (a generalisation of the partial trace in quantum theory) as follows. For some state s of a multi-partite system, the reduced state s_A is defined such that it encodes *all* the statistics of s that could be measured locally at A only.

2.6.1 The tensor product for states and for effects

As discussed earlier, the tensor product concerns itself with the composition of subsystems into a larger whole. The effect written $\mathbf{e}_{a_i} \otimes \mathbf{e}_{b_j}$ represents the functional acting on a bipartite state mapping it to the joint probability of receiving outcome a_i for some measurement A that can only be made by Alice, and outcome b_j for some measurement B that can only be made by Bob; $P(a_i, b_j | A, B) = (\mathbf{e}_{a_i} \otimes \mathbf{e}_{b_j}) \cdot \mathbf{s}$ for some bipartite state $\mathbf{s}$. As the unit effect simply returns the normalisation of the state-space, if Alice makes a measurement, but is not interested in what Bob measures, then her effects are of the form $\{\mathbf{e}_{a_i} \otimes \mathbf{1}\}_i$.

Similarly, we can combine states using this notation. Expressing the state of a multi-partite system as a tensor product of states of subsystems indicates a lack of correlation between the subsystems. Suppose initially, there is a system $\mathbf{s}_{AB}$ in which there is absolutely no correlation between measurements made on the constituent parts A and B . This means that the output of any measurement $\mathbf{e}_A$ on the system A does not change as a result of learning the outcome $\mathbf{e}_B$ of a measurement on the system B . This allows us to factorise the probabilities of a joint measurement on the combined system into the two measurements on the reduced states of the constituent systems:

$$P(\mathbf{e}_A, \mathbf{e}_B | \mathbf{s}_A, \mathbf{s}_B) = P(\mathbf{e}_A | \mathbf{s}_A) P(\mathbf{e}_B | \mathbf{s}_B) = (\mathbf{e}_B \cdot \mathbf{s}_A) (\mathbf{e}_B \cdot \mathbf{s}_B). \quad (2.36)$$

and express the joint state of $\mathbf{s}_A$ with $\mathbf{s}_B$ using the tensor product symbol $\otimes$:

$$\mathbf{s}_{AB} = \mathbf{s}_A \otimes \mathbf{s}_B, \quad (2.37)$$

and so along with the joint effect on A and B

$$\mathbf{e}_{AB} = \mathbf{e}_A \otimes \mathbf{e}_B, \quad (2.38)$$

we write

$$\mathbf{e}_{AB} \cdot \mathbf{s}_{AB} = (\mathbf{e}_A \otimes \mathbf{e}_B) \cdot (\mathbf{s}_A \otimes \mathbf{s}_B) = (\mathbf{e}_A \cdot \mathbf{s}_A) (\mathbf{e}_B \cdot \mathbf{s}_B), \quad (2.39)$$

so that the probabilities factorise as in eq. (2.36).

We can extend this notation to transformations. The tensor product $T_A \otimes T_B$ expresses a transformation T_A on A and T_B on B applied simultaneously in such a way that no correlations between A and B are generated. That is

$$(\mathbf{e}_A \otimes \mathbf{e}_B) \cdot (T_A \otimes T_B) (\mathbf{s}_A \otimes \mathbf{s}_B) = (\mathbf{e}_A \cdot T_A \mathbf{s}_A) (\mathbf{e}_B \cdot T_B \mathbf{s}_B). \quad (2.40)$$

Finally, we make a remark that some theories have a property known as being *locally tomographic*.

Definition 2.11: Locally tomographic

A theory is **locally tomographic** if the state of the whole system can always be completely determined by the measurements made at local parts, and the correlations between these locally-possible measurements.

This is not a trivial restriction (see *Principle 5* in [16]). For example in ‘real quantum mechanics’, which is quantum mechanics on a real Hilbert space, the Pauli σ_y matrix is not a valid measurement, as it is a complex matrix, but the correlation $\sigma_y \otimes \sigma_y$ between two systems is allowed, because it is a real matrix. Thus, the theory is not locally-tomographic, and there are properties of the joint system that are not correlations between properties of the constituent parts.

2.6.2 Example: P-R boxes

The no-signalling restriction (described above) does not limit our joint states-spaces to quantum theory. It is possible to construct states which are more non-local⁸ than are admitted in the quantum state-space, but do not violate no-signalling [54, 60, 61]. The states that violate non-locality by the highest amount possible are commonly referred to as P-R (Popescu–Rohrlich) boxes, and are part of the *box-world* framework (the associated single system state-spaces are gbits). Such systems provide a natural foil to quantum non-locality, sharing many of the same consequences [62] (e.g. no-cloning and no-broadcasting theorems [63, 64]). As such, the foundational physics community has dedicated a lot of attention to the study of these systems (see [55, 65] and references within).

An example of a P-R box (for binary measurements A_1 and A_2 made by Alice, and B_1 and B_2 made by Bob, each with outcomes $+$ and $-$) could be written:

⁸ Non-locality is the concept that measurements may correlate in a way that is neither due to exchange of a signal, nor due to prior arrangement between the parties that they will measure in a particular way [55]. The usual quantification of this property for bipartite systems is the CHSH-Bell inequality [53, 56–59].

$$s_{\text{PR}} = \left(\begin{array}{c} P(+, +|A_1, B_1) \\ P(+, -|A_1, B_1) \\ P(-, +|A_1, B_1) \\ P(-, -|A_1, B_1) \\ \hline P(+, +|A_1, B_2) \\ P(+, -|A_1, B_2) \\ P(-, +|A_1, B_2) \\ P(-, -|A_1, B_2) \\ \hline P(+, +|A_2, B_1) \\ P(+, -|A_2, B_1) \\ P(-, +|A_2, B_1) \\ P(-, -|A_2, B_1) \\ \hline P(+, +|A_2, B_2) \\ P(+, -|A_2, B_2) \\ P(-, +|A_2, B_2) \\ P(-, -|A_2, B_2) \end{array} \right) = \left(\begin{array}{c} \frac{1}{2} \\ 0 \\ 0 \\ \frac{1}{2} \\ \hline \frac{1}{2} \\ 0 \\ 0 \\ \frac{1}{2} \\ \hline \frac{1}{2} \\ 0 \\ 0 \\ \frac{1}{2} \\ \hline 0 \\ \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{array} \right). \quad (2.41)$$

For such a state, the outcomes from measurements A_1 and B_1 , A_1 and B_2 , A_2 and B_1 are all correlated; whereas the outcomes from measurements A_2 and B_2 are anti-correlated. This means that the CHSH equation [57] takes its numerical maximum:

$$\langle A_1, B_1 \rangle + \langle A_1, B_2 \rangle + \langle A_2, B_1 \rangle - \langle A_2, B_2 \rangle = 4. \quad (2.42)$$

No such measurements could be found that satisfy this property for *any* joint quantum state.

On the face of it, P-R boxes appear to have great computational power. The existence of P-R boxes would allow for distributed computing with *trivial computational complexity*: the calculation of a function $f(A, B) = \{0, 1\}$, where A is a string known only to Alice and B is a string known only to Bob, can be done by only sharing a single bit of communication between Alice and Bob, if they have access to shared P-R boxes [66–68] (cf. quantum theory which requires a number of bits at least equal to the length of the strings [67, 69], even if Alice and Bob share arbitrarily many entangled states).

On the other hand, the reversible dynamics available for P-R boxes (and box-world in general) are very restricted, in that only permutations of measurements, measurement outcomes, and subsystem labels are permitted [50]. In particular, this means there are no correlating interactions, and so no reversible transformation exists that could transform gbits into P-R boxes and vice versa⁹. Likewise, limitations on

⁹ If an analogous result of this strength held in quantum theory, it would prevent the preparation of maximally entangled Bell states from product states.

the allowed box-world measurements prevent the implementation of entanglement swapping, teleportation or dense coding [70, 71] (all of which would be possible in quantum theory).

2.7 Spekkens' toy model

In the final section of this chapter, we shall introduce a toy model, first posed by Robert Spekkens, that exhibits behaviour close to quantum behaviour in certain scenarios [28, 29]. (This model is not a priori part of the GPT framework, but we will show how it can be extended so that it is.)

Spekkens initially posed this toy model in consideration of whether quantum states should be viewed as *ontic* (a state of reality), or *epistemic* (a state of knowledge). The model demonstrates that a lot of non-classical behaviour can be recovered by placing *epistemic restrictions* (that is, a limit on what can be known) on a system that might have some underlying ontic state, and hence was presented as evidence in favour of an epistemic view of quantum theory (cf. [72] arguing in favour of an ontic interpretation). In particular, states in Spekkens' toy model are restricted by the principle that the amount of knowledge we have about the ontic state is exactly the same as the amount of knowledge that is not known.

The simplest system in the toy model consists of a classical system which admits 4 possible ontic states, outlined in table 2.1.

Numerical label	'Square' diagram	'Flat' diagram
1		
2		
3		
4		

Table 2.1: Single bit ontic states of Spekkens' toy model. *Of the two diagrammatic notations, we remark that the 'flat' notation is more commonly found in the literature, as it can be extended naturally into multi-partite systems. In the parts of this thesis that are predominantly concerned with the behaviour of single systems, we will use the square notation which takes up less space when used in line.*

None of the states in table 2.1 are allowed to represent our knowledge about the system, as this would be in violation of the epistemic restriction: choosing 1 from 4 amounts to two bits of data, and there would be nothing left to know about the system.

To satisfy the epistemic restriction, a state of knowledge may narrow down the system to at most 2 of the 4 possibilities.

In this elementary system (which we call a *Spekkens' bit*) selecting 2 possibilities can be viewed as the answer to a binary question, for instance (visualising in the concise square notation) “is the ontic state of the system in the top of the square?”. If the answer to the above is yes, then the underlying state is either  or , and we express our state of knowledge (the *epistemic state*) as . On the other hand, if the answer was no, then the ontic state would either be  or , and we would express the epistemic state as .

Combinatorics tells us that there are six binary questions we can ask about the ontic state of the form “is it in A or B , as opposed to in C or D ?” As we can assign a unique epistemic state to each positive answer of such questions, it is convenient to label the different epistemic states by the choice of question. Moreover, these six questions divide in to three pairs, where in each pair one and only one question can give a positive answer. Thus in direct analogy with how states of a qubit may be labelled by the ± 1 eigenvalues of measurement by a Pauli matrix, we can label the three pairs of questions as X , Y and Z ; and for each pair, choose one possibility to be $+1$ and the other -1 . The choices we shall use in this text are given by table 2.2, and are consistent with Table II of [73].

Measurement and outcome	Supported ontic states	Square diagram	Flat diagram
$X = +1$	$1 \vee 3$		
$X = -1$	$2 \vee 4$		
$Y = +1$	$1 \vee 4$		
$Y = -1$	$2 \vee 3$		
$Z = +1$	$1 \vee 2$		
$Z = -1$	$3 \vee 4$		
Maximally mixed	$1 \vee 2 \vee 3 \vee 4$		

Table 2.2: Single bit epistemic states of Spekkens' toy model. $a \vee b$, the union of a and b , is read to mean “exactly one of a or b ”.

Although all six of the questions have definite answers given an ontic state, the same questions could be asked about epistemic states. It is obvious that an epistemic state  should answer positively to the question “is it ?” all the time, and likewise should always answer negatively to “is it ?” However, if the question was “is it

“ $\begin{smallmatrix} \blacksquare & \square \\ \square & \blacksquare \end{smallmatrix}$?”), it would a priori be impossible to answer without knowing precisely whether the ontic state was $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ or $\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}$, which is forbidden. Naïvely, making the measurement (taken together with prior knowledge that the state was in one of these possibilities) would allow us to totally specify the state.

In this toy model, however, we postulate that all states which we assign the same epistemic state are totally indistinguishable, regardless of whether they might have an underlying ontic state. It is this postulate (above even the epistemic restriction) that makes the toy model an epistemic model rather than ontic one: it is an express declaration that our knowledge about the system is of paramount operational importance.

Thus, when we ask “is it $\begin{smallmatrix} \blacksquare & \square \\ \square & \blacksquare \end{smallmatrix}$?” on an epistemic state $\begin{smallmatrix} \blacksquare & \square \\ \square & \square \end{smallmatrix}$, we can no longer demand a definite answer, but rather instead accept a probabilistic one: the answer is equally likely to be yes or no. (In general, the probability is given by the ratio of ontic states in the epistemic state that overlap with those being considered in the measurement). Furthermore, once the answer has been given, we must change the state, such that it is now in a new epistemic state ($\begin{smallmatrix} \blacksquare & \square \\ \square & \square \end{smallmatrix}$ if positive, or $\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}$ if negative), to ensure that if the same question is asked twice in a row, the answers are consistent.

If we wish to maintain the concept of an underlying ontic state at all (which is effectively functioning as a hidden variable in this model), we thus introduce an update rule that, following a measurement, the ontic state is randomly mapped to one of any of the possibilities consistent with post-measurement epistemic state.

This behaviour is very similar to how complementary measurements on a qubit work. Suppose a quantum state $|0\rangle$ was measured by $X = |+\rangle\langle+| - |-\rangle\langle-|$. With probability 1/2, the output would be +1 and the post measurement state $|+\rangle$, and with probability 1/2, the output would be -1 and the post measurement state $|-\rangle$. Exactly the same is true for the state $|1\rangle$ being measured by X . Thus, one could make a measurement in the $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$ basis, followed by one in the X , followed by a third in the Z again; and with 50% probability, the two Z measurements will give contradictory answers. This same scenario happens in Spekkens toy model when one measures $\begin{smallmatrix} \blacksquare & \square \\ \square & \square \end{smallmatrix}$ vs $\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}$, followed by $\begin{smallmatrix} \blacksquare & \square \\ \square & \square \end{smallmatrix}$ vs $\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}$, followed by another measurement of $\begin{smallmatrix} \blacksquare & \square \\ \square & \square \end{smallmatrix}$ vs $\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}$.

Finally, we remark that the state where *nothing* is known about a system is also permitted. This state, the maximally mixed state, is written $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ or $\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}$.

2.7.1 Transformations

The allowed transformations in Spekkens’ toy model correspond to the maps on the underlying ontic state that do not take valid epistemic states to invalid states. To

calculate the action of a transformation on an epistemic state, one takes the map of each ontic possibility supported within that state in turn, and combines them.

Many-to-one maps are not permissible, as acting on a valid epistemic state might result in an over-specified invalid state. For instance, if both $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ and $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ are mapped to $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$, then the valid epistemic state $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ will be mapped to the invalid state $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$.

One-to-many maps are permissible so long as each ontic possibility occurs in the output state the same number of times (i.e. so that in the new epistemic state, every supported ontic possibility is equally likely). For instance, if every ontic state is mapped to $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ or $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ with equal probability; this corresponds to an operation that irreversibly prepares the state $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$.

Of prime interest however are the reversible one-to-one maps, which are given by the permutations on the ontic state (group S_4). For the single bit systems, every permutation is valid, in that every valid epistemic state will be taken to a valid epistemic state. It is sometimes convenient to label these permutations using the numerical value of the ontic states, and use the one-row notation for describing permutations. In this notation $wxyz$ represents the permutation that replaces ontic possibility 1 with w , 2 with x , etc. where w , x , y and z are unique choices of the elements 1, 2, 3 and 4. It is also sometimes convenient to express the permutations graphically, as arrows linking the initial and final ontic state (see fig. 2.6).

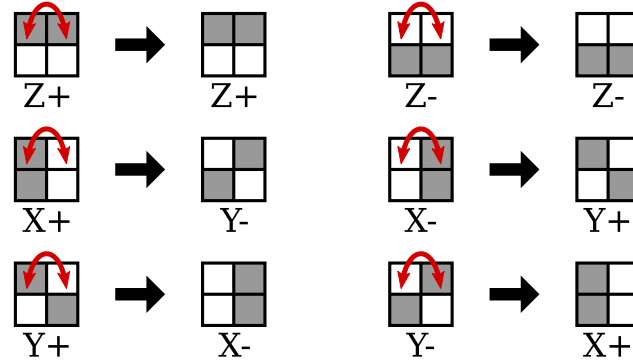


Figure 2.6: Example gate in Spekkens' toy model. The permutation 2134 which exchanges $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ and $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ is applied to the six 'pure' epistemic states.

2.7.2 Multi-partite systems

A bipartite system composed of two single bit systems contains 16 ontic configurations in total (4×4). We can represent these 16 ontic possibilities on a 4×4 grid, where each axis corresponds to the ontic possibilities of one system. For instance the ontic state

where system A is in $\begin{smallmatrix} \blacksquare & \square \\ \square & \square \end{smallmatrix}$ but system B is in $\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}$ would correspond to the co-ordinate $(3, 1)$ in this grid (we will sometimes write this as $\begin{smallmatrix} \blacksquare & \square \\ \square & \square \end{smallmatrix} \otimes \begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}$).

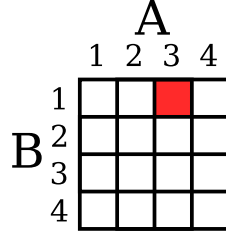


Figure 2.7: Bipartite state-space in Spekkens' toy model The ontic state where A is in ontic possibility $\begin{smallmatrix} \blacksquare & \square \\ \square & \square \end{smallmatrix}$ and B in ontic possibility $\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}$ is highlighted in red.

The 16 possibilities correspond to 4 bits of data, but the epistemic restriction of the model only allows us to know at most 2 bits of information about the state. The purest bipartite epistemic states hence consist of just 4 ontic possibilities of the joint system. However, unlike for a single-bit system, we cannot just pick 4 possibilities arbitrarily, as the bipartite structure places additional requirements: namely that the epistemic restriction must also apply on single system subspaces within the joint state.

A multi-partite system can be 'reduced' by flattening one axis such that all states that only differ by their co-ordinate in that axis correspond to the same state in the reduced system. For some epistemic states, this could potentially map multiple supported ontic states onto just one ontic in the reduced system. This is allowed, as long as each ontic possibility in the reduced system corresponds to the same number of supported ontic states in the full system. If this were not the case, measurements on a reduced system would no longer satisfy the rule that every supported state contributes equally to the probability of a particular outcome. Compliance with this rule limits which choice of ontic possibilities can constitute a valid multi-partite epistemic state.

Moreover, reduction to a single axis is not allowed to result in an epistemic state that itself violates the epistemic principle. In the context of the 4×4 grid for bipartite system, this rules out states where the four possibilities lie in a single column or row.

Finally, and more subtly, we do not permit states where the outcome of measurement on one system could over-specify the state on another. For instance the example $\begin{smallmatrix} \blacksquare & \square \\ \square & \square \end{smallmatrix} \otimes \begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix} + \begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \square \\ \square & \square \end{smallmatrix}$, supports 4 possible ontic states (see Eq. 56 in [28], also illustrated in fig. 2.8). If the Z measurement were made on the first system, then the exact ontic state of the second becomes known. The subtlety of why this is a problem this lies in how we define the post-measurement update rule on joint systems: the 'shuffling' that updates the state of the joint system following a local measurement is itself a locally applied permutation (such that the way in which a joint ontic state is changed does not depend on the ontic state of the other (not measured) system); see caption of fig. 2.8.

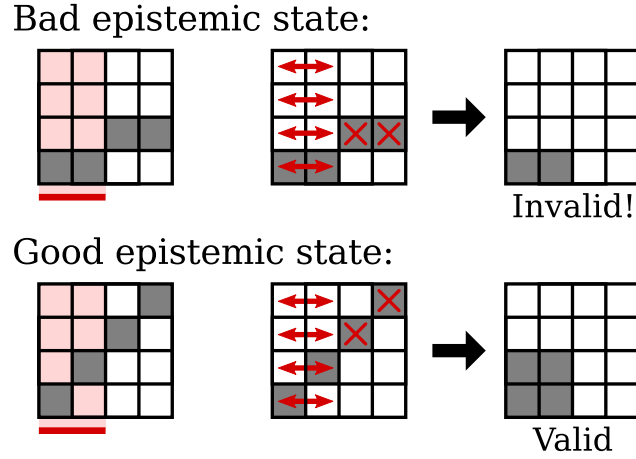


Figure 2.8: Good and bad epistemic states. The Z measurement is made on bit A (see fig. 2.7), and the $Z+$ outcome obtained (highlighted in pink on the left-most grid). The update rule means we discard out the unsupported ontic possibilities (marked with red crosses), and then apply a random permutation in A on the remaining possibilities (marked with arrows). For the top row this results in a clearly invalid state, which has too few ontic possibilities. The bottom row, however, is fine.

With all these restrictions, we are limited to three classes of bipartite states (shown in fig. 2.9): **i. Product states**, where each of the parties has maximal knowledge of their own system, and any measurement made by one party does not allow for any additional information about the outcome of measurements made by the other party. **ii. Pure correlated states**, where the reduced state for both parties is maximally mixed; but any measurement made by one party allows that party to predict the corresponding outcome for a measurement by the other party. **iii. Mixed correlated states**, where for just one measurement made by one party, an associated measurement by the other party can be predicted.

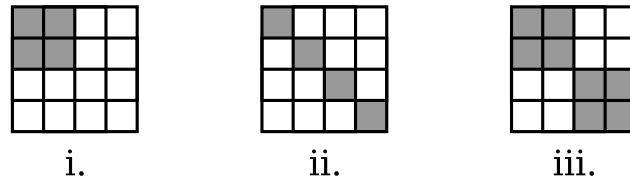


Figure 2.9: Example allowed bipartite epistemic states. i) Product states; ii) Pure correlated states; iii) Mixed correlated states. Full description in main text.

Combinations of product states where one (or both) sides attempt to combine non-orthogonal (or non-identical) states are not allowed. For instance, $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} + \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ is not allowed, as the reduced state for either party would be the forbidden state $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ (in which the ontic possibility $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ has twice the weighting as $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ or $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$).

2.7.3 Convex extension

Canonically, Spekkens' toy model only admits epistemic states where the probability of any particular ontic state being occupied is either 0 or a non-zero value that is the same for every supported ontic state in that epistemic state. It is sufficient to know if an ontic possibility is supported or not; and the probability of the system being in that state can be determined directly from the number of possibilities. (This allows us to draw epistemic states as black and white diagrams).

In the GPT framework, a state of knowledge about a classical system with N possibilities could be described by a N -dimensional real vector, whose elements are in the range $[0, 1]$. Recall that the normalised state-spaces of classical systems correspond to simplices (§ 2.1), and so the state-space of a four-state system is the tetrahedron. We could associate each ontic state in the elementary single-bit system of Spekkens' model with one of these vertices, such that the possible state of the system is now described by an arbitrary probability distribution. (This leads to a construction similar to [74]). Continuing with this picture, the six valid epistemic states (table 2.2), in which two of possibilities occur with equiprobability, can be associated with the midpoint of each edge between the supported possibilities. The maximally mixed state  lies at the centre $s = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

To draw this 4-level state-space in three dimensions, we impose the restriction that every state is normalised (reducing our 4 degrees of freedom to 3). Moreover, because the positive and negative outcomes of the three different measurements correspond to mutually exclusive ontic possibilities, the epistemic states corresponding to the two outcomes of each measurement are on opposite edges, sharing no vertices. For the tetrahedron, each edge only has one non-connected opposite edge. We can hence orient the state-space such that states associated with positive and negative outcomes of each measurement X , Y and Z are aligned orthogonally, and so are drawn in alignment with the Cartesian axes of the diagram (see fig. 2.10).

The natural extension of Spekkens' toy model into the convex framework is to admit the convex hull of the canonically allowed states (i.e. the smallest convex set which contains all the allowed epistemic states). By doing this on the elementary system, we end up with an octahedral state-space, as drawn in the RHS of fig. 2.10. This state-space is strictly a subset of the probability distributions for a normalised 4-level classical system, which is the tetrahedral state-space as drawn on the LHS of fig. 2.10.

Equivalently (for the single-bit system), one could place the requirement that we do not admit states where the probability of the system being in any particular ontic possibility is greater than 0.5. Writing a choice of the four ontic states as X_1, X_2, X_3 and X_4 such that $X_1 \neq X_2 \neq X_3 \neq X_4$ and $X_1, X_2, X_3, X_4 \in \left\{ \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}, \begin{smallmatrix} \blacksquare & \square \\ \blacksquare & \square \end{smallmatrix}, \begin{smallmatrix} \square & \blacksquare \\ \square & \blacksquare \end{smallmatrix}, \begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix} \right\}$, we have 4 different inequalities of the form $P(X_1) \leq 0.5$, corresponding to the facets formed by

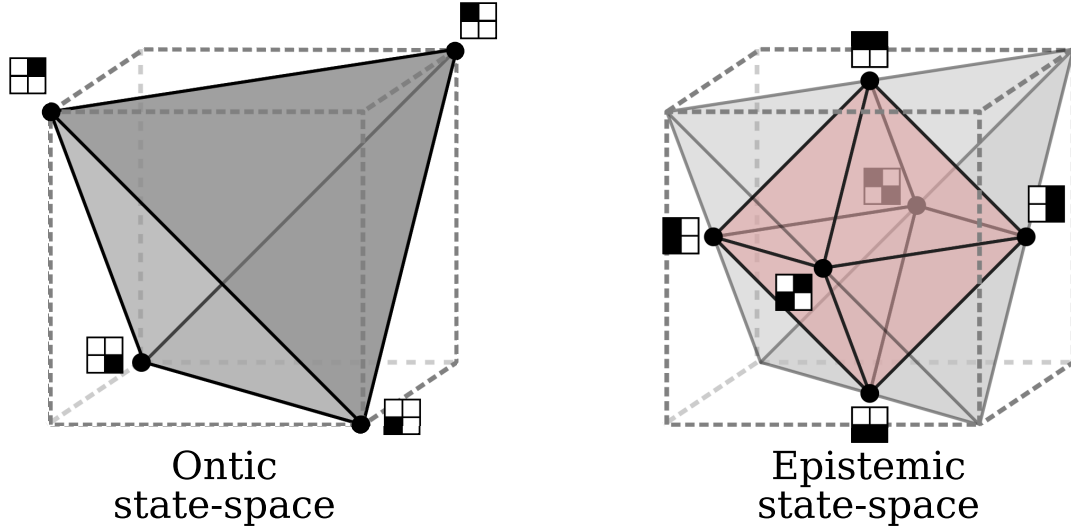


Figure 2.10: State-space of single bit in Spekkens' toy model. The left-hand side shows the tetrahedral state-space for the Spekkens bit without epistemic restriction. The right-hand side shows the octahedral epistemic state-space, consisting only of the states which satisfy the epistemic restriction. The RHS state-space is a strict subset of the LHS state-space.

'cutting off' a corner of the tetrahedron. (The other 4 facets, which exist also in the tetrahedron, consist of the set of normalised states that exclude one possibility, such that $P(X_1) \geq 0$, which satisfy $P(X_2) + P(X_3) + P(X_4) = 1$.)

We refer to the octahedral state-space as the **epistemic state-space** of the theory, as it only admits states compatible with the epistemic restriction. On the other hand, the full set of conceivable ontic states we call the **ontic state-space** (though we note that in both cases, the state could be a probabilistic object!)

Drawing the state-spaces in this way invites comparison to other theories in the GPT framework, in particular with 3-in 2-out gbits (i.e. cubic state-spaces), and the standard quantum qubit, which both also have three complementary measurements. For the tetrahedron, the four pure states correspond to states where the outcome of all three measurements could be simultaneously predicted. These states violate the uncertainty principle, and are hence not allowed states for a qubit, but are all valid gbit states (and moreover are pure gbit states). Not every pure gbit state is realisable in the ontic state-space, however. The tetrahedron has only 4 corners, and so 4 out of the 8 possible gbit pure states are not allowed: the tetrahedron only admits pure states that pairwise share one measurement outcome¹⁰.

¹⁰There are only two choices of which subset of pure gbit states are admitted in the tetrahedron. Visually, this corresponds to there only being two distinct ways of drawing a tetrahedron whose vertices are shared with the corners with a cube; a rotation of the

On the other hand, the epistemic state-space is completely contained within the Bloch sphere of quantum mechanics; and so every valid epistemic state could be prepared in quantum theory. In fact, this octahedral state-space is exactly the same as the state-space of stabilizers in quantum theory [73], and so the state-space corresponds to a limited subset of valid qubit states.

This comparison with the stabilizer subset of quantum mechanics breaks down when it comes to the allowed transformations [28]. This is because the octahedral state-space suggests more symmetry than is actually allowed. In fact, the reversible transformations are those which map the tetrahedral state-space to itself, and so are the S_4 subgroup of the full octahedral symmetries. The full octahedral automorphism group given by $S_4 \times Z_2$ admits operations such as the 90° proper rotation (also admissible in quantum theory), that would map the tetrahedron onto its non-overlapping mirror image. If such an operation were allowed on the epistemic state-space of Spekkens' toy model, it would correspond to an action on the epistemic states that could not be induced by a permutation of the ontic possibilities. For this reason, we rule it out. The implications of this transformation behaviour will be discussed further in §3.3.4.

The simple extension of the toy model outlined in this section (mirroring that presented in [1]) is sufficient for the purposes for which the toy model will be employed within this thesis, where we will be predominantly concerned with the behaviour of pure states under transformation. Other extensions of Spekkens' toy model into the GPT framework may be found by van Enk [74], and by Janotta and Lal [75].

tetrahedron by 90° about any of the three measurement axes, or reflecting any co-ordinate, will result in the same (alternative) shape.

Chapter 3

Phase in the GPT framework

At the heart of quantum theory lies the concept of *phase*. Phase is a key component of quantum computing: many quantum algorithms may be expressed as phase estimation procedures [26, 76]. There also exists a class of quantum algorithms that consist only of operations that alter the phases between (non-entangled) inputs in the computational basis (multiple copies of $|0\rangle$ vs. $|1\rangle$), which are argued to have stronger computational power than classical computers [27, 77]. Phase is also the core concept underlying various physical phenomena such as interferometry (which we shall elaborate upon in the subsequent chapters of this thesis), thermalisation (e.g. [78]) and decoherence (e.g. [79]).

What if phase and its related phenomena were not uniquely quantum concepts? At first glance, the idea of phase appears to be intrinsically tied with arguments of complex numbers in Hilbert spaces. In order to study phase in non-quantum theories (where we might no longer be able to use the well-defined mathematical structure of Hilbert spaces), we must first lay down the foundations for its study in a generalised context. In this chapter¹, we shall do this for the GPT framework.

First, we shall revisit the mathematics of phase in quantum theory (§3.1). We will then introduce a compatible definition of phase, suitable for use in the GPT framework, phrased in terms of the dynamics of the system with respect to a chosen measurement (§3.2). We will show how this definition leads to a well-defined mathematical structure. We will provide reversible (§3.3) and irreversible (§3.4) examples of this within the GPT framework. We shall then remark on the definition with respect to a special case of measurements known as *maximal measurements* (§3.5), and how this allows us to highlight phase as a signifier for non-classicality. We shall also prove a few lemmata linking the phase dynamics of related measurements (§3.6). Finally, taking quantum POVMs as an example, we will show how the definition and derived results of this chapter can be applied to a general class of measurements (§3.7).

¹Some of the results in this chapter were published in [1].

3.1 Phase in quantum systems

Before suggesting a generalisation of the concept in GPT framework, we will first briefly review the mathematics of phase in quantum theory.

3.1.1 Phase in pure quantum states

Consider an orthonormal basis $\{|j\rangle\}$ associated with outcomes of a (projective) measurement M . In this basis, an arbitrary pure normalised quantum state is written

$$|\psi\rangle = \sum_j c_j e^{i\phi_j} |j\rangle, \quad (3.1)$$

where $\{c_j\}$ and $\{\phi_j\}$ are real numbers, and $\sum_j c_j^2 = 1$. The probability of measurement M on state $|\psi\rangle$ yielding outcome j is given by

$$P(M = j) = \langle j|\psi\rangle\langle\psi|j\rangle = |c_j|^2. \quad (3.2)$$

These outcome probabilities assigned to measurement M are independent of the choices of $\{\phi_j\}$. In quantum parlance, these terms $\{\phi_j\}$ are known as the *phases* associated with each outcome.

In any basis multiplying the entire state by a factor of $e^{i\theta}$ will not change the outcome probabilities of *any* measurement made on that state (following from $|e^{i\theta}|^2 = 1$). This *global phase* has no measurable effect, and only the relative phases between outcomes are operationally important. The self-duality of Hilbert spaces allows every pure quantum state to be associated with an outcome that occurs with certainty of some measurement. Thus, for every pure state, there will always be a basis where the state effectively has no phases whatsoever (except for the unphysical global phase), illustrating that it is meaningless to talk about the phase of a state without reference to a choice of basis².

Consider now a unitary transformation of the form

$$U = \sum_k e^{i\xi_k} |k\rangle\langle k| \quad (3.3)$$

expressed also in the basis associated with M . The effect of this unitary operation is

²In a lot of contexts, this choice of basis might be implicit when one measurement is of obvious importance; such the branches of an interferometer, or the energy levels of a two-level system.

to take $|\psi\rangle$ to $|\tilde{\psi}\rangle$ such that

$$|\tilde{\psi}\rangle = U|\psi\rangle = \sum_j c_j e^{i(\phi_j + \xi_j)} |j\rangle. \quad (3.4)$$

The outcome probabilities for measurement M on state $|\tilde{\psi}\rangle$ are given by exactly the same expression as for $|\psi\rangle$, $P(M = j) = |c_j|^2$. The only effect of such a U to change the phases $\phi_j \rightarrow \phi_j + \xi_j$, and for this reason is known as a *phase transformation*³ with respect to measurement M .

The set of phases for all states with particular probabilities for one measurement is isomorphic to the set of phase transformations, as can be seen by choosing a reference state $|\psi_0\rangle = \sum_j c_j |j\rangle$ (for real numbers c_j , and all $\phi_j = 0$). By writing

$$|\psi\rangle = U|\psi_0\rangle = \sum_j c_j e^{i\xi_j} |j\rangle, \quad (3.5)$$

we see that through choice of $\{\xi_j\}$ in U , we cover all possible phases in the state $|\psi\rangle$.

Finally, we note that the set of all $\{U\}$ forms a group of reversible transformations under composition, which in quantum theory is $U(1) \oplus U(1) \oplus \dots \oplus U(1) \subset U(N)$ where N is the dimension of the Hilbert space. This group naturally arises in the context of an interferometer as the deterministic operations on the disjoint branches of the interferometer, which can adjust the relative path lengths, but not cause the traversing particle to jump from one branch to another (discussed further in chapter 4.)

3.1.2 Phase in general quantum states

Consider now a general N -dimensional quantum state ρ composed of complex elements $\{\rho_{ij}\}$. The probability of projective measurement M yielding outcome j (associated with state $|j\rangle\langle j|$) is given by

$$P(M = j) = \text{Tr}(\rho |j\rangle\langle j|) = |\rho_{jj}|^2. \quad (3.6)$$

The off-diagonal terms $\{\rho_{ij}\}$ where $i \neq j$ do not affect the probabilities of any of the outcomes associated with this measurement.

For pure states (in the form eq. (3.1)), one sees for the associated density matrix that $\rho_{ij} = c_i c_j \exp(i(\phi_i - \phi_j))$. The argument of ρ_{ij} gives the relative phase between the two outcomes i and j . For mixed states, the phase between possibilities is not always well-defined. In particular, if there is total decoherence between two possibilities, then

³Sometimes referred to as a *phase operation* in contexts when ‘phase transformation’ may ambiguously refer to changes between states of matter (e.g. solid to liquid).

the corresponding density matrix element ρ_{ij} is 0, and so the argument (and hence relative phase) is undefined.

Talking about *phase transformations* avoids this problem. In the context of general quantum states, these are the set of allowed quantum operations (i.e. a subset of completely positive maps) that do not alter the statistics associated with measurement M . That is, for every outcome j associated with projector $|j\rangle\langle j|$ in the measurement M , we can consider maps $\Phi : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$ that take $\rho \rightarrow \sigma$ such that

$$P(M = j|\rho) = \text{Tr}(|j\rangle\langle j|\rho) = \text{Tr}(|j\rangle\langle j|\Phi(\rho)) = P(M = j|\sigma) \quad \forall \rho, j \quad (3.7)$$

to be *phase transformations* associated with measurement M .

3.2 Defining phase in the GPT framework

We can draw on the above notion of quantum phase transformations to formulate a generalisation of the concept in the GPT framework.

Definition 3.1: Phase transformation

For a system in a GPT with at least one measurement M (composed of the set of effects $\{\mathbf{m}_i\}$), a **phase transformation** T_ϕ associated with measurement M is a valid transformation on the system that does not alter any of the statistics associated with measurement M for any allowed state $\mathbf{s}$ in the system:

$$P(M = m_i|\mathbf{s}) = \mathbf{m}_i \cdot \mathbf{s} = \mathbf{m}_i \cdot T_\phi \mathbf{s} = P(M = m_i|T_\phi \mathbf{s}) \quad \forall \mathbf{s}, i. \quad (3.8)$$

This definition is an immediate generalisation of eq. (3.7) into the GPT framework. In the quantum case, the states are density matrices, the effects are projectors, the probabilities are calculated using the Hilbert-Schmidt inner product, and the allowed transformations are a subset of completely positive maps. Moreover, this definition implies that reversible phase transformations in quantum theory have the usual $\bigoplus \text{U}(1)$ form:

Lemma 3.1. *For an N -level quantum system, with a projective measurement Z , whose outcomes may be expressed in terms of projectors $\{Z_j = |j\rangle\langle j|\}_{j=1\dots N}$ (and $\sum_j Z_j = \mathbb{1}$), all reversible phase transformations (according to definition 3.1) may be written in the form*

$$U = \sum_j e^{i\phi_j} |j\rangle\langle j|. \quad (3.9)$$

Conversely, all matrices of this form are a phase transformation of Z according to definition 3.1.

Proof. Using the isomorphism between density matrices and states, and effects and measurement operators, the condition for a transformation U to be a phase transformation associated with Z (eq. (3.8) in definition 3.1) can be written in the quantum formalism as:

$$\text{Tr}(U\rho U^\dagger Z_j) = \text{Tr}(\rho Z_j) \quad \forall \rho, j. \quad (3.10)$$

This implies that $[U, Z_j] = 0$ for all j , and so we can write U diagonally in the basis of Z ,

$$U = \sum_j e^{i\phi_j} |j\rangle\langle j|. \quad (3.11)$$

To show the converse, because U of the above form commutes with all elements Z_j , and $U^\dagger U = \mathbb{1}$, we may use the cyclicity of trace to write: $\text{Tr}(Z_j U \rho U^\dagger) = \text{Tr}(U^\dagger Z_j U \rho) = \text{Tr}(U^\dagger U Z_j \rho) = \text{Tr}(Z_j \rho)$, which is exactly the phase condition written in eq. (3.10). $\square$

(We shall later prove a similar result in lemma 5.3, for the general form of *irreversible* quantum phase transformations.)

3.2.1 The structure of phase transformations

It will prove useful to talk about the set of phase transformations related to a particular measurement, and so we define the following term in the context of the GPT framework:

Definition 3.2: Phase dynamics of a measurement

The set of all allowed phase transformations associated with a particular measurement in a theory is known as the **phase dynamics** of that measurement. This is the set of all allowed T_Φ that satisfy eq. (3.8) for that choice of measurement.

The caveat ‘allowed’ in the above definition accommodates for the possibility of linear transformations satisfying definition 3.1 that are forbidden by a theory for some other reason⁴. For example, reflections in the XZ plane of a Bloch sphere for a qubit in quantum theory would preserve the statistics of a Z measurement (satisfying eq. (3.8)), but correspond to a map that is not completely positive, and hence are disallowed by

⁴Recall the distinction between *valid* and *allowed* transformations, discussed in §2.3.5: valid transformations do not violate any rule of the GPT framework, allowed transformations are valid transformations that are also not ruled out by any theory-specific restriction.

quantum theory. Such transformations thus not be considered part of a theory's phase dynamics, as they are not a part of its dynamics in general.

We can show that the set of phase dynamics, together with the operation of composition (that is, applying one transformation in the set followed by another in the set) form a monoidal structure.

Lemma 3.2 (Monoidal structure of phase dynamics). *The set of phase transformations and the operation composition form a monoid (a semigroup with an identity element).*

Proof. The proof follows by demonstrating the following properties required for a set and operation to form a *monoid*:

- i. Closure,
 - ii. Associativity,
 - iii. Existence of an identity element.
- i. Let T_A and T_B both be arbitrary phase transformations (satisfying eq. (3.8)) for some measurement M . Suppose we act first with T_A on s to form s' with the same M statistics as s . As s' is an allowed state, it is in the domain of T_B , and so we can act s' with T_B to form s'' , a state with the same M statistics as s' and hence the same M statistics as s . The same logic will follow for any choice of state s . Thus the composition of phase transformations T_A and T_B is itself a transformation that satisfies eq. (3.8), and so is a phase transformation, demonstrating *closure*.
 - ii. Recall that all linear transformations on vectors may be written as real-valued matrices, and the composition of such transformations is matrix multiplication. The associativity of elements within the set of phase dynamics thus follows from the associativity of matrix multiplication (see, e.g. §7.3.2 in [80]). (I.e. phase dynamics are associative under composition because all GPT transformations are associative under composition.)
 - iii. Consider the trivial transformation, $T = \mathbb{1}$ which does not change any statistics whatsoever. This automatically satisfies $\mathbf{m}_i \cdot s = \mathbf{m}_i \cdot \mathbb{1}s$ because $\mathbb{1}s = s$, and so 3.8 holds trivially, and $\mathbb{1}$ is itself a phase transformation.

By satisfying i, ii, and iii, the set of phase transformations forms a monoidal structure. □

We can go one step further, and arrive at a *group* structure if we restrict ourselves to reversible transformations (see definition 2.7). Noting that the set of all reversible transformations forms a group under composition, we see that the restriction in eq. (3.8), induces a subgroup of this.

Lemma 3.3 (Group structure of reversible phase dynamics). *The set of all allowed reversible phase transformations and the operation of composition form a group.*

Proof. A group must satisfy closure, associativity, the existence of an identity element, and the existence of an inverse element for every element within it. The first three of these are proven in lemma 3.2, so let us now consider the final property regarding the existence of inverse elements.

Noting that the general set of reversible transformations forms a group (the automorphism group of the state-space, or a subgroup thereof), we must prove that if T satisfies eq. (3.8) for some measurement M , then the inverse T^{-1} defined such that $TT^{-1} = \mathbb{1}$ must also satisfy eq. (3.8) for that measurement.

First consider a phase transformation $T : s \rightarrow s' = Ts$, such that s and s' are valid states sharing the same M statistics. Note that T^{-1} reverses the action of T such that $T^{-1} : s' \rightarrow s$. For every s' that can be formed by acting with T on some s , T^{-1} hence acts like a phase transformation because s' and s share the same M statistics. However, the image of a state-space through a reversible transformation is the state-space itself ($\{s\} = \{s'\}$), and so when T satisfies the phase transformation requirement for all s , T^{-1} satisfies it for all s' , and hence for every allowed state. This argument is made for arbitrary T ; hence every reversible transformation T which is a phase transformation has a T^{-1} which is also a phase transformation.

We conclude that the subset of reversible transformations induced by eq. (3.8) is a group. $\square$

We call this group the phase group of a measurement:

Definition 3.3: Phase group of a measurement

The **phase group** associated with a measurement M on a system is the subset of all allowed reversible transformations that are also phase transformations, with the group operation of composition.

Before we move onto a few examples, we remark on a final piece of structure that holds for the set of phase dynamics under the probabilistic mixing of transformations.

Lemma 3.4 (Convexity of phase dynamics). *Any probabilistic mixture of phase transformations for a particular measurement is also a phase transformation for that measurement.*

Proof. Consider a measurement M with effects $\{\mathbf{m}_i\}$. Let T_A and T_B be two arbitrary phase transformations of M (satisfying eq. (3.8)). Consider a convex mixture of these transformations $T = \alpha T_A + (1 - \alpha) T_B$ for $\alpha \in [0, 1]$.

From the linearity of this mixing, $T : s \rightarrow s'$ is given by:

$$s' = Ts = \alpha T_A s + (1 - \alpha) T_B s. \quad (3.12)$$

Consider evaluating the effect $\mathbf{m}_i$ on the initial state s :

$$\mathbf{m}_i \cdot s = (1 - \alpha + \alpha) \mathbf{m}_i \cdot s = (1 - \alpha) \mathbf{m}_i \cdot s + \alpha \mathbf{m}_i \cdot s. \quad (3.13)$$

As T_A and T_B are both phase transformations of M , we can substitute $\mathbf{m}_i \cdot s = \mathbf{m}_i \cdot T_A s$ and $\mathbf{m}_i \cdot s = \mathbf{m}_i \cdot T_B s$, and write

$$\mathbf{m}_i \cdot s = (1 - \alpha) \mathbf{m}_i \cdot T_A s + \alpha \mathbf{m}_i \cdot T_B s, \quad (3.14)$$

and then use linearity (of matrix multiplication and the inner product) to factor out $\mathbf{m}_i$ on the left hand side, and s from the right hand side:

$$\mathbf{m}_i \cdot s = \mathbf{m}_i \cdot ((1 - \alpha) T_A + \alpha T_B) s = \mathbf{m}_i \cdot Ts. \quad (3.15)$$

Hence, $\mathbf{m}_i \cdot s = \mathbf{m}_i \cdot Ts$, and this holds for all $\mathbf{m}_i$, all s and all choices of α . This shows that mixing phase transformations always results in another phase transformation. $\square$

3.3 Examples of phase groups

We will discuss how the above definitions apply to the reversible dynamics of a few theories within the GPT framework.

3.3.1 Classical probability

The state of a classical system can be described by the outcome statistics of a single measurement. The only valid phase transformation associated with such a measurement is the identity operator, which changes nothing about the system at all. For this reason, the phase group $\{\mathbb{1}\}$ associated with such a measurement is said to be *trivial*.

However, for classical systems larger than just a binary degree of freedom, it is possible to find measurements that do not fully specify the system, and hence allow non-trivial phase dynamics (a concept we will discuss further in § 3.5). An example can be found in a system of two classical bits, with binary measurements M_1 and M_2 measuring the state of the first and second bit respectively. The group of local transformations consists of combinations of flipping (X) or leaving unchanged ($\mathbb{1}$) each bit: $Z_2 \oplus Z_2 = \{\mathbb{1} \oplus \mathbb{1}, \mathbb{1} \oplus X, X \oplus \mathbb{1}, X \oplus X\}$, and (amongst other global operations)

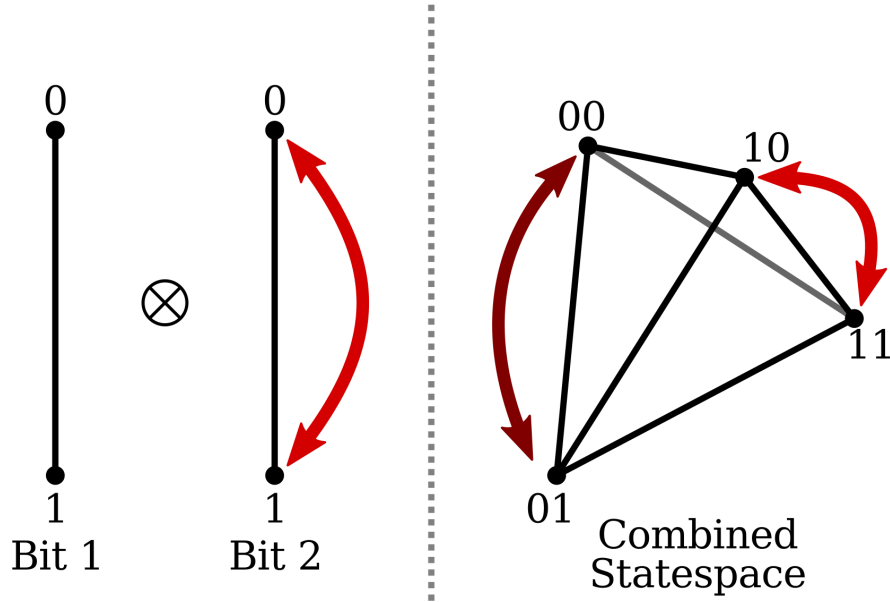


Figure 3.1: Phase group associated with measurement of one bit in a system of two classical bits. The LHS shows the phase group $\mathbb{1} \oplus \mathbb{Z}_2$ associated with local operations: flipping the second bit, or leaving it alone. The RHS shows the global set of operations in this phase group. When neither arrow applies, nothing changes; when both apply, the second bit is unconditionally flipped; when only the left (right) applies, the second bit is flipped conditionally on the state of the first bit being 0 (1).

there are also four distinct global correlating interactions that trigger a X -flip on one of the bits conditional on the state of the other. The phase group associated with measurement M_1 is the group of operations that only change the statistics associated with M_2 ; the group $\{\mathbb{1} \oplus \mathbb{1}, \mathbb{1} \oplus X, C_{Z_1}X_2, C_{\bar{Z}_1}X_2\}$ (where the final two elements are classical CNOT gates that flip the second bit if the first bit is in state 1 or state 0 respectively). This group is illustrated in fig. 3.1.

3.3.2 Quantum theory: qubits

Recall from § 2.4.2 that the state-space of a qubit can be represented by the Bloch sphere, and the allowed reversible transformations are given by the rotation group $SO(3)$. Consider a binary measurement with mutually exclusive outcomes made on a qubit. These outcomes correspond to antipodes of the sphere, and the line between them defines an axis. Restricting transformations to the phase group associated with this measurement implies that the projection onto this axis must not be changed.

This creates an invariant axis of rotation⁵, and so induces the phase-group $SO(2)$: the rotations of a circle (illustrated in Figure 3.2). $SO(2)$ is isomorphic to $U(1)$, the phase group that would be predicted by the quantum definition in eq. (3.3) for a two-level system, once the global phase has been taken into account.

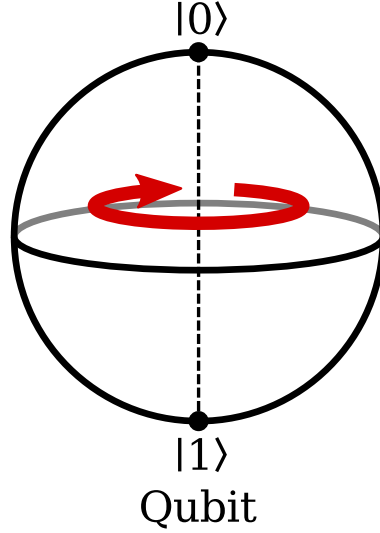


Figure 3.2: Phase group associated with Z measurement of a qubit: $SO(2)$ rotations around the Z -axis of the Bloch sphere. No rotation about an axis will change the value of the projection of the state onto the axis.

3.3.3 Box-world: 3-in 2-out gbts

Consider a cubic state-space of 3 binary measurements X , Y and Z , each with outcomes labelled $+1$ and -1 , with no restrictions on the statistics. Writing the state as a column vector of probabilities in the form given by eq. (2.11) in §2.3.2, we can consider the effects associated with the outcomes of the Z measurement:

$$\mathbf{e}_{Z+} = (0, 0 | 0, 0 | 1, 0), \quad \mathbf{e}_{Z-} = (0, 0 | 0, 0 | 0, 1). \quad (3.16)$$

For a state $\mathbf{s}$, one notes that any transformation T that only affects the statistics of X and Y measurements will not change the values of $\mathbf{e}_{Z+} \cdot \mathbf{s}$ and $\mathbf{e}_{Z-} \cdot \mathbf{s}$. For normalised binary measurements the probability associated with one outcome is sufficient to determine all the statistics, and so this leaves two degrees of freedom remaining: one for X and one for Y . Without further constraint, this corresponds to the symmetries

⁵In general, the subgroup of $SO(N)$ that leaves one axis invariant is $SO(N-1)$ [81].

of a square. For this reason, the phase group of a fiducial measurement in a 3-in 2-out gbit is the dihedral group $\mathcal{D}_4$ (illustrated in the left-hand side of fig. 3.3).

However, this is not the only type of binary measurement that could be made on the system. Consider if instead we chose a more complicated measurement M defined by its effects:

$$\mathbf{e}_{M+} = \frac{1}{3} (1, 0 | 1, 0 | 1, 0), \quad \mathbf{e}_{M-} = \frac{1}{3} (0, 1 | 0, 1 | 0, 1). \quad (3.17)$$

As discussed in § 2.4.3, these effects generate the statistics that would occur if one measured X , Y or Z with equal probability, recorded the outcome, but not which of X , Y or Z was measured. Thus, the group of reversible transformations that preserve the statistics of M correspond to permutations of measurements as a whole (e.g. swapping the X and Y statistics). Hence, the phase group associated with measurement M is the permutation group S_3 , giving a significantly different set of transformations than the phase group associated with one of the fiducial measurements. This second group is illustrated on the right-hand side of fig. 3.3.

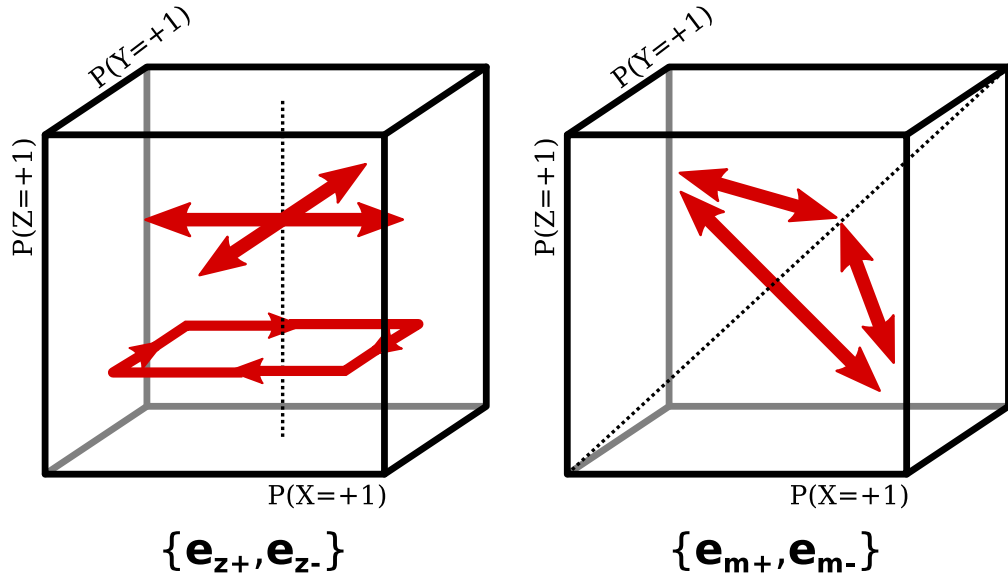


Figure 3.3: Two classes of phase groups for a single 3-in 2-out gbit. The LHS shows the phase group of a Z measurement: the dihedral group of order 2, corresponding to the automorphisms of a square (reflections and 90° rotations). The RHS shows the phase group of the measurement defined by effects $\mathbf{e}_{M+}$ and $\mathbf{e}_{M-}$, generating the permutation group allowing three-way rotations (and flips) about the diagonal axis from state $(1, 0 | 1, 0 | 1, 0)^T$ to $(0, 1 | 0, 1 | 0, 1)^T$.

Both these types of measurements generalise naturally to higher-dimension hypercubic gbts (n -in 2-out). The phase group associated with a single measurement $\mathbf{e}_{z+}$ vs. $\mathbf{e}_{z-}$ will be the automorphisms of an $(n-1)$ -dimensional hypercube. The phase group associated with the mixed measurement (generalising $\mathbf{e}_{M+}$ and $\mathbf{e}_{M-}$ to be the

effects associated with equiprobably choosing one of the n measurements, recording the outcome, but not the measurement) will be the simplex group of order n , S_n .

We remark that for m -in 2-out qbits where $m \geq 2$, these phase groups are non-abelian; the order in which the phase transformations are applied has a measurable effect on the output statistics. The potential physical consequences of this will be discussed in chapter 6.

3.3.4 Spekkens' toy model

Recall from §2.7.1 that the allowed transformations in Spekkens' toy model correspond to permutations of the ontic state, and hence the automorphisms of a tetrahedron (§2.7.3). If we freeze the Z measurement (associated with  vs. ), then we obtain the subgroup of permutations $\{1234, 2134, 1243, 2143\}$: $Z_2 \oplus Z_2$. The four elements of this group correspond to operations that exchange the top two corners ($2134 = X \oplus \mathbb{1}$), exchange the bottom two corners ($1243 = \mathbb{1} \oplus X$), exchange both ($2143 = X \oplus X$), or exchange neither ($1234 = \mathbb{1} \oplus \mathbb{1}$). Such a group is illustrated in the left hand side of Figure 3.4.

This group is exactly the same as the group obtained by fixing the value of one of two classical bits (described in §3.3.1). This is because the hidden variable describing the ontic state of the model is, in fact, two classical bits. One of these bits, corresponding to the Z measurement in Spekkens' model, is fixed, and the second bit corresponding to (e.g.) the X measurement) is free to change its value. The 2143 and 1234 operations locally flip (or preserve) the value of the second bit, regardless of the value of the first. The permutations 2134 and 1243 create correlations between the two classical bits by only flipping the X value if the Z value is in a particular state.

We remark that this group structure agrees with the allowed operations for Spekkens' toy model, as described by Coecke et al. [82]. The *phase group* as introduced here, and the phase group in their paper are not a priori connected, being defined in very different mathematical frameworks. (For instance, their definition requires the phase group to be abelian, whereas the one presented here does not). At this stage, the precise connection between these two definitions has not been firmly established. However, as both the results presented here and the results of [82] are attempting to capture some property of the dynamical behaviour of a system, it is mutually reassuring that the predictions are the same⁶.

In the manner of the qbit example in §3.3.3, we could also consider a more exotic

⁶We leave it for the reader to prove that the phase group (by our definition) of a measurement on single system stabilizer system (which has the same state-space as the epistemic state-space of Spekkens' toy model [73]) is Z_4 , in agreement with the result in [82].

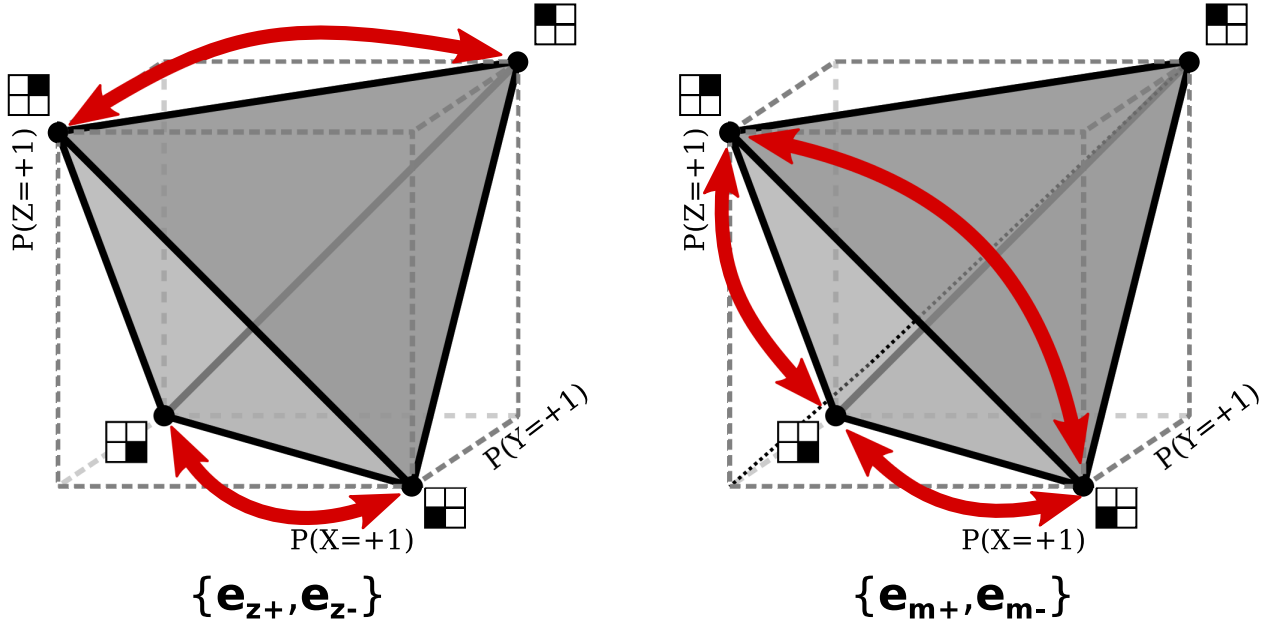


Figure 3.4: Phase groups for a single bit in Spekkens' toy model (shown on the ontic state-space). The LHS shows the phase-group of the Z measurement: the pair of flips $Z_2 \oplus Z_2$. The RHS shows the phase group of the mixed measurement e_{M+} vs. e_{M-} : the permutation group S_3 about the diagonal axis between the center of one of the triangular faces, and the opposite point.

measurement, associated with effects

$$e_{M+} = \frac{1}{3} (1, 0 \mid 1, 0 \mid 1, 0), \quad e_{M-} = \frac{1}{3} (0, 1 \mid 0, 1 \mid 0, 1). \quad (3.18)$$

Unlike the gbit, there are no states that give a certain outcome when taking the dot product with these effects. Rather, the most extremal points for e_{M+} are on the face that mixes $\left\{ \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \right\}$, and for e_{M-} are on the face that mixes $\left\{ \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \right\}$. The statespace has three planes of reflection that do not mix between these sets, as well as three-fold rotational symmetry around the diagonal axis defined by e_{M+} and e_{M-} . These are the symmetries of a triangle, and thus we see again that this measurement has the associated phase group S_3 .

Finally, we note that for the (forbidden) measurement that distinguishes between the four ontic states $\left(\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}, \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}, \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}, \text{ and } \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \right)$, the associated phase group is trivial (consisting only of the identity), as this measurement completely determines the system, and so there is nothing that can be changed (this concept will be formalised later in theorem 3.5).

3.4 Examples of irreversible phase dynamics

In the examples given in § 3.3, we considered reversible phase operations (phase groups). The admitted phase groups varied dramatically according to the theory being considered. However, it is also interesting to consider irreversible phase dynamics, and note that these too can differ between theories.

3.4.1 Decoherence

A qubit with X , Y and Z measurements (outcomes ± 1) has a general state vector:

$$\mathbf{s} = \begin{pmatrix} P(X = +1) \\ P(X = -1) \\ \frac{P(Y = +1)}{P(Y = -1)} \\ \frac{P(Z = +1)}{P(Z = -1)} \end{pmatrix}. \quad (3.19)$$

Consider the quantum operation associated with the Pauli matrix σ_z . The effect of this operation is a 180° rotation about the Z -axis in the Bloch sphere, exchanging the two X eigenstates (and also exchanging the two Y eigenstates). In the real-valued representation of eq. (3.23), this operation may be written as the (block-diagonal) matrix:

$$T_{\sigma_z} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \mathbb{1}_2 \quad (3.20)$$

Suppose, with equal probability, we either apply this operation or do nothing ($\mathbb{1}_6$). In the GPT framework, the net effect of such an operation is given by convex combination $T_D = \frac{1}{2}T_{\sigma_z} + \frac{1}{2}\mathbb{1}_6$:

$$T_D = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \oplus \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \oplus \mathbb{1}_2. \quad (3.21)$$

Applying this operation to a general (normalised) quantum state of the form eq. (3.23) gives (recalling that $P(X = +1) + P(X = -1) = 1$, and likewise for Y):

$$T_D : \mathbf{s} \rightarrow \mathbf{s}_D = T_D \mathbf{s} = \begin{pmatrix} 1/2 \\ 1/2 \\ \frac{1/2}{1/2} \\ \frac{1/2}{1/2} \\ \frac{P(Z = +1)}{P(Z = -1)} \\ \frac{P(Z = -1)}{P(Z = -1)} \end{pmatrix}. \quad (3.22)$$

Because the statistics for the Z measurement are the same for $\mathbf{s}$ and $\mathbf{s}_D$ for any

choice of $\mathbf{s}$, T_D is a phase transformation with respect to the Z measurement: $\mathbf{e}_{Z+} \cdot \mathbf{s} = \mathbf{e}_{Z+} \cdot T_D \mathbf{s}$ and $\mathbf{e}_{Z-} \cdot \mathbf{s} = \mathbf{e}_{Z-} \cdot T_D \mathbf{s}$ for all choices of $\mathbf{s}$ (using the same definitions of $\mathbf{e}_{Z\pm}$ as in § 3.3.3). This is as we would expect from lemma 3.4, as both $\mathbb{1}_6$ and T_{σ_Z} are phase transformations of the Z measurement.

This operation is exactly that of perfect *decoherence*: as it completely randomises the outcomes of the X and Y measurements. Consider for example the pure state $\mathbf{s}_{X+}$ corresponding to the $+1$ eigenstate of the X measurement $|+\rangle$, and the maximally mixed state $\mathbf{s}_{\text{mm}}$:

$$\mathbf{s}_{X+} = \begin{pmatrix} 1 \\ 0 \\ \frac{1/2}{1/2} \\ \frac{1/2}{1/2} \\ \frac{1/2}{1/2} \end{pmatrix}, \quad \mathbf{s}_{\text{mm}} = \begin{pmatrix} 1/2 \\ 1/2 \\ \frac{1/2}{1/2} \\ \frac{1/2}{1/2} \\ \frac{1/2}{1/2} \end{pmatrix}. \quad (3.23)$$

In this example, $T_D \mathbf{s}_{X+} = \mathbf{s}_{\text{mm}}$. The operation is irreversible because the same map would also take the state $\mathbf{s}_{X-}$ with opposite X statistics to the same maximally mixed state $\mathbf{s}_{\text{mm}}$. (In fact, all states in that lie in the plane of the equator of the Bloch sphere would be mapped to $\mathbf{s}_{\text{mm}}$).

As remarked upon in § 2.3.5, a transformation matrix having a determinant of 0 is sufficient⁷, but not necessary, for an operation to be irreversible. Even within the example of quantum decoherence, one could have a weighted mixture of T_{σ_Z} and $\mathbb{1}$ for the partial decoherence transformation: $T_{PD} = (1 - \epsilon)\mathbb{1}_6 + \epsilon T_{\sigma_Z}$ for $\epsilon \in (0, 1)$. The effect of this (for $\epsilon \neq \frac{1}{2}$) is to shrink the state-space from the Bloch sphere to an ellipsoid keeping the Z -axis unchanged (see fig. 3.5). Such a transformation is irreversible because the inverse matrix T_{PD}^{-1} would map allowed states that are not inside the ellipsoid to a point outside the Bloch sphere.

3.4.2 ‘Measurement setting’ in a gbit

3-in 2-out gbts admit irreversible phase operations that are not possible in quantum theory. One such example is the transformation

$$T_{\text{set}X} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \oplus \mathbb{1}_4, \quad (3.24)$$

which sets the X statistics of any state to $(1, 0)$. This is a phase transformation of the Z (and also of the Y) measurements, as these statistics are unaltered. Such an trans-

⁷In a geometric context, the determinant T_D is obviously zero because the image of the 3-dimensional Bloch sphere is collapsed to a volumeless 1-dimensional line (the Z -axis).

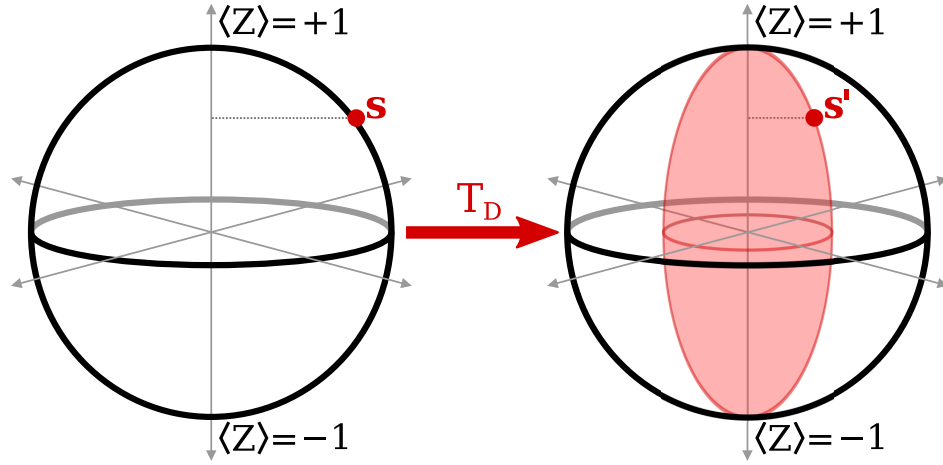


Figure 3.5: Partial decoherence transformation on a qubit. The full state-space on the LHS is mapped into the pink ellipsoid on the RHS by the partial decoherence transformation T_D . For any state s (e.g. the red dot on the LHS) mapped onto s' (red dot on RHS), the projection onto the Z axis remains unchanged, and so T_D is a phase transformation.

formation is referred to as a ‘measurement setting’ operation, as it unconditionally overrides the statistics of one measurement.

This operation is not valid in quantum theory, as it maps allowed states to disallowed states. For example, the quantum state $|0\rangle$ written as $\mathbf{s}_{Z+} = (1/2, 1/2 | 1/2, 1/2 | 1, 0)^T$ transforms into $T_{\text{set}X} \mathbf{s}_{Z+} = (1, 0 | 1/2, 1/2 | 1, 0)^T$, which is not a valid quantum state as it violates the uncertainty principle by allowing perfect knowledge of the X and Z simultaneously.

A similar, but valid, quantum operation is

$$T_{\text{prep}X} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \oplus \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \quad (3.25)$$

This operation changes any state to $(1, 0 | 1/2, 1/2 | 1/2, 1/2)^T$ (i.e. $|+\rangle$). Unlike $T_{\text{set}X}$, the preparation transformation $T_{\text{prep}X}$ is *not* a phase transformation for the Z measurement, as there are some allowed input states (e.g. $|0\rangle$) whose Z statistics are not equiprobable, and so will be altered by the transformation.

3.5 Phase transformations of *maximal measurements*

Consider a two qubit system, with the binary measurement $M = Z \otimes \mathbb{1}$. According to definition 3.1, the phase transformations associated with this measurement include

Z -preserving transformations on the first qubit, but also include *arbitrary* operations on the second qubit. This allows a much more permissive set of operations than would be allowed by the phase group associated with a four-outcome measurement (e.g. measuring the output in the Z basis for both qubits).

In a more extreme example, which holds for all theories in the GPT framework, the phase group of the measurement whose only effect is the unit effect 1 contains every normalisation-preserving transformation. Such a groups may appear to violate the spirit of phase by being too inclusive. Whilst we have carefully provided a definition of phase in terms of any measurement, we now propose a particular type of measurement, whose related phase-dynamics are likely to be more characteristic in any given theory.

Definition 3.4: Maximal measurement

A measurement is described as a **maximal measurement** in a theory if it can always unambiguously distinguish between the maximum number of states possibly distinguishable by a single measurement in that theory.

In the language of [9], these are basis measurements: they distinguish a number of mutually exclusive outcomes equal to the *dimension* of the system. The caveat ‘always’ is to exclude certain measurements where there is a probability of an outcomes that in fact yields no new information about the state (as will be discussed in § 3.7.1). The caveat ‘unambiguously’ means that for each effect e_i in the measurement, there is at least one state s_i such that $e_i \cdot s_i = 1$ and $e_j \cdot s_i = 0$ for all other effects in the measurement. Such a measurement would thus be able to discriminate between $\{s_i\}$, without chance of misidentification. If there are no other sets of effects in the theory that can distinguish between more than N such states, then M is a maximal measurement.

The two examples at the beginning of this section (as well as the two-bit classical example in § 3.3.1) are not maximal measurements, as they only distinguish between two outcomes in theories where one could provide a measurement that distinguishes perfectly between four cases. These examples are all *coarse-grained* measurements. Coarse-grained measurements are never maximal by their very construction: they are formed by combining two or more effects of a finer measurement into a single outcome in the new coarser measurement. This coarser measurement cannot distinguish as many different states as the finer measurement could, and so is not maximal.

For the ontic states of a bit in Spekkens’ toy model, the maximal measurement distinguishes between four possibilities, and only allows the trivial phase operation (just like a four-outcome classical system). However, such a measurement is ruled out in the usual (epistemic) Spekkens’ toy model, where a measurement can only distinguish

perfectly between two possibilities. We see that phase dynamics- a somewhat ‘non-classical’ behaviour- has been recovered by imposing a restriction on knowledge. This result is very much in line with the original conception of the toy model, illustrating how incomplete knowledge can accommodate deviation from classical behaviour [28]. By this example, we see that the existence of phase dynamics does not rule out hidden variable models.

Even restricting to maximal measurements is not sufficient to determine a unique phase group for a particular theory. Recall the gbit with three binary measurements: e_{Z+} vs. e_{Z-} can distinguish perfectly between states where $P(Z = 1) = 1$ (any state from the top plane of the cube) and states where $P(Z = 1) = 0$ (a state from the bottom plane of the cube). By making the measurement e_{M+} vs. e_{M-} , one can perfectly distinguish between gbit states $s_+ = (1, 0|1, 0|1, 0)^T$ and $s_- = (0, 1|0, 1|0, 1)^T$. There are no measurements that can distinguish in a single trial between more than two states, and so both these measurements are maximal, even though they are associated with different phase groups.

In quantum theory, however, there is only one type of phase group for all maximal measurements on a system of a particular dimension. No quantum measurement can distinguish with certainty between more possibilities than the dimension of the Hilbert space. Each projective quantum measurement can be transformed into any other by a unitary operation. Thus, for a phase group $\{\sum_j e^{i\phi_j} |j\rangle\langle j|\}$, a unitary transformation converts each $|j\rangle\langle j| \rightarrow U|j\rangle\langle j|U^\dagger = |\tilde{j}\rangle\langle \tilde{j}|$ giving the phase group $\{\sum_j e^{i\phi_j} |\tilde{j}\rangle\langle \tilde{j}|\}$. These two groups are isomorphic to each other (though they preserve different axes), and are both representations of $\bigoplus^N U(1)$ where N is the Hilbert space dimension.

3.5.1 Phase and non-classicality

Using the concept of maximal measurements, we can make a couple of statements that link phase with the concept of non-classicality.

Definition 3.5: Non-classical

A theory is **non-classical** if it admits maximal measurements, whose statistics do not in general uniquely specify the state of the system.

(This definition is consistent with non-classicality being a discrepancy between the *dimension* and the number of *degrees of freedom* of a state, as described by [9]). Conversely, a system is **classical** if its state can be totally specified by the outcome statistics of any maximal measurement on the system.

Theorem 3.5. *The phase dynamics associated with a maximal measurement on a*

classical system are always trivial⁸.

Proof. A classical system can be completely determined by the outcome of one measurement. As there is nothing else to be known about the state, there is only one transformation that does not disturb these statistics: $\mathbb{1}$. Thus the phase dynamics of a maximal measurement on a classical system are trivial. $\square$

We thus see it is *necessary* for the system to be non-classical to have non-trivial phase dynamics. Phrasing the converse is less straight-forward, as there could be arbitrary conditions in a theory forbidding various transformations. One construction was phrased in §4.2 of [1], but this relies on the existence of a decoherence operation (e.g. § 3.4.1) with respect to at least one measurement in the theory. Even requiring that the set of transformations in a theory form a convex set is not sufficient to argue for the existence of an allowed decoherence operation. For example, consider a highly asymmetrical state-space whose automorphism group is $\{\mathbb{1}\}$; mixtures of $\{\mathbb{1}\}$ are also trivial and so to have non-trivial dynamics, the theory would need to admit some transformations beyond the convex hull of its reversible transformations.

Without losing generality, we thus make the following (more cautious) statement:

Lemma 3.6. *A non-classical system always has at least one valid phase transformation associated with a maximal measurement (although this valid transformation may be disallowed by some other restriction on the dynamics of the theory).*

Proof. The proof follows by showing that if there is more information in a system than encoded in a single maximal measurement, then it is possible to conceive of a decoherence map that takes valid states to valid states.

Consider a maximal measurement M (one which determines the most possible information about the system; see definition 3.4). Let the $\{t_i\}$ be a set of states unambiguously distinguished by M with associated effects e_i , such that $e_i \cdot t_j = \delta_{ij}$.

We explicitly construct a map that projects the state-space onto the hull of $\{t_i\}$, by considering the function that prepares one of the state t_i with probability given by the effect of e_i acting on an arbitrary input state s , $e_i \cdot s$:

$$Ts = s' = \sum_i (e_i \cdot s) t_i. \quad (3.26)$$

The input of this function is an allowed state s ; and as $\sum_i e_i \cdot s = \mathcal{N}$, and $e_i \cdot s \in [0, \mathcal{N}]$, the output is also guaranteed to be a valid state in the state-space (it is in the convex hull of known allowed states $\{t_i\}$). Such an operation is also manifestly linear in its input s .

⁸Definition 2.8.

Moreover, T is a phase operation. Consider a state s where $P(M = i|s) = e_i \cdot s$. The state after projection is given $s' = \sum_i (e_i \cdot s) t_i$, and thus we can evaluate $e_j \cdot s'$ as $\sum_i (e_i \cdot s) e_j \cdot t_i = e_j \cdot s = P(M = j|s)$ (where we have used $e_i \cdot t_j = \delta_{ij}$, as M unambiguously and perfectly distinguishes the set $\{t_i\}$). This is true for all e_j and so T is a phase operation with respect to M .

To show that T is not trivial, we use the fact that the state of a non-classical system cannot be determined by the outcome of a single measurement. It may be that for some (or all) of t_i , knowing the statistics of the measurement is enough to completely specify the state (e.g. knowing $P(Z = +1)$ specifies a unique state $|0\rangle$ in quantum theory), but there must be at least one configuration of M -statistics where there is ambiguity in the choice of state (at least for one choice of maximal measurement); as if this were not the case, the system would be classical as it could be specified by the statistics of a single measurement.

Labelling two such states with shared M statistics as s_1 and s_2 , because state-spaces are convex, any probabilistic combination $s = \lambda s_1 + (1 - \lambda) s_2$ is also an admitted state. Hence s_1 and s_2 can be found such that for just one $\lambda \in [0, 1]$, the state s is in the convex hull of t_i . At least one of s_1 or s_2 will not be in the hull of $\{t_i\}$ and so will have been transformed in a non-trivial way by T .

Thus, for non-classical theories, it is possible to find a non-trivial valid phase transformation associated with a maximal measurement. $\square$

3.6 The phase dynamics of related measurements

The conditions in definition 3.1 for a transformation to be a phase transformation form a set of linear constraints. These constraints induce a subset of the complete dynamics: the phase dynamics (definition 3.2). Although so far, we have defined phase dynamics with respect to a set of effects that form a measurement; as linear constraints, the mathematics of the following lemmata will follow for arbitrary sets of effects $\{e_i\}$ even when $\sum_i e_i \neq 1$. Although more general sets of effects lose their physical interpretation as measurement outcomes (as they will in general fail to assign probability distributions to states), these lemmata will still prove helpful mathematical statements to relate measurements with shared features. This will allow us to later prove statements about specific kinds of phase groups (such as POVMs in §3.7).

We begin by formalising the intuitive concept that adding extra constraints on the allowed transformations can only ever reduce the set of possible phase dynamics.

Lemma 3.7. *Consider two sets of effects $\mathcal{E} = \{e_i\}$ and $\mathcal{F} = \{f_j\}$ acting on the same system, such that the phase dynamics induced from eq. (3.8) are $S_{\mathcal{E}}$ and $S_{\mathcal{F}}$ respec-*

tively. The union of these sets of effects $\mathcal{EF} = \mathcal{E} \cup \mathcal{F}$ has phase dynamics given by the intersection of the individual sets of dynamics $S_{\mathcal{EF}} = S_{\mathcal{E}} \cap S_{\mathcal{F}}$.

Proof. This follows as a straightforward consequence how linear constraints are combined, elaborated on below for completeness.

Let T be a transformation in the ‘phase’ dynamics associated with the set of effects $\mathcal{E} \cup \mathcal{F}$. The constraints in $\mathcal{E} \cup \mathcal{F}$ mean that T satisfies $\mathbf{e}_i \cdot \mathbf{s} = \mathbf{e}_i \cdot T\mathbf{s}$ for all $\mathbf{e}_i$ and $\mathbf{s}$ and hence T is an element of $S_{\mathcal{E}}$. Moreover, T also satisfies $\mathbf{f}_i \cdot \mathbf{s} = \mathbf{f}_i \cdot T\mathbf{s}$ for all $\mathbf{f}_i$ and $\mathbf{s}$ and so T is also an element of $S_{\mathcal{F}}$. As T is an element of both $S_{\mathcal{E}}$ and $S_{\mathcal{F}}$, it is by definition in the intersection $S_{\mathcal{EF}} = S_{\mathcal{E}} \cap S_{\mathcal{F}}$.

Suppose there was some T both in $S_{\mathcal{E}}$ and in $S_{\mathcal{F}}$ (and hence in the intersection $S_{\mathcal{EF}}$) that was not a phase dynamic associated with $\mathcal{EF}$. As T satisfies all the constraints of $\mathcal{E}$ and $\mathcal{F}$ separately, then we would require an additional constraint in $\mathcal{EF}$ that explicitly rules it out, and so $\mathcal{EF} \neq \mathcal{E} \cup \mathcal{F}$ leading to contradiction. $\square$

The linearity of the phase constraints allows us to make statements relating the sets of phase dynamics between linearly related sets of effects.

Lemma 3.8. *If a set of effects $\{\mathbf{m}_j\}$ can be expressed as a linear combination of effects $\{\mathbf{n}_i\}$, then the respective sets of phase dynamics S_M and S_N satisfy $S_N \subseteq S_M$. If the mapping between $\{\mathbf{m}_j\}$ and $\{\mathbf{n}_i\}$ is reversible, then equality holds such that $S_N = S_M$.*

Proof. When $\{\mathbf{m}_j\}$ is expressed linearly in terms of effects of $\{\mathbf{n}_i\}$, we can write

$$\mathbf{m}_j = \sum_i c_{ji} \mathbf{n}_i \quad (3.27)$$

for every j , where $\{c_{ji}\}$ is a set of constant (real) scalar values.

Let T be an allowed phase transformation with respect to $\{\mathbf{n}_i\}$, such that $\mathbf{n}_i \cdot \mathbf{s} = \mathbf{n}_i \cdot (T\mathbf{s})$. Multiplying both sides by c_{ji} , then summing over all i yields:

$$\begin{aligned} \sum_i c_{ji} \mathbf{n}_i \cdot \mathbf{s} &= \sum_i c_{ji} \mathbf{n}_i \cdot (T\mathbf{s}), \\ \mathbf{m}_j \cdot \mathbf{s} &= \mathbf{m}_j \cdot (T\mathbf{s}), \end{aligned} \quad (3.28)$$

where the second line follows from the linearity of the inner product. A similar argument can be made for all j , and hence T also satisfies the criteria to be a phase transformation of $\{\mathbf{m}_j\}$. Every transformation T in S_N is also in S_M , and so $S_N \subseteq S_M$.

Suppose that there is a real inverse of eq. (3.27), such that there exists a set of coefficients $\{\tilde{c}_{ji}\}$ that allow the relationship $\mathbf{n}_j = \sum_i \tilde{c}_{ji} \mathbf{m}_i$. The same argument as above can be made to show that $S_M \subseteq S_N$. Taken simultaneously with $S_N \subseteq S_M$, this implies that $S_M = S_N$ when there is a reversible mapping between $\{\mathbf{m}_j\}$ and $\{\mathbf{n}_i\}$. $\square$

From the mathematical lemma 3.8, we can make a few straight-forward corollaries that apply to more physical measurements:

Corollary 3.9. *Coarse-graining a measurement does not reduce the set of phase dynamics.*

Proof. If measurement M is a coarse-graining of measurement N , then each effect in M can be expressed as a sum of effects in N . Hence, the first claim of lemma 3.8 is applicable, and every transformation allowed by N is also allowed by M . $\square$

Corollary 3.10. *Relabelling measurement outcomes⁹ does not change the associated phase dynamics.*

Proof. If M is a relabelling of N , then their effects are related by a permutation which is inherently invertible. Hence, the second claim of lemma 3.8 is applicable, and the phase dynamics will thus be the same. $\square$

We see this lemma and its corollaries predict a set structure consistent with the two extremal cases of measurements possible on a system. A measurement that reveals *everything* about the system (e.g. a maximum measurement on a classical system, or a general informationally complete measurement, as will be discussed in § 3.7.2) has only the trivial phase group $\{1\}$, and this element is included in the phase dynamics of *every* measurement. At the other extreme, a measurement that reveals *nothing* about the system (e.g. a single outcome measurement associated with the unit effect 1) admits every allowed normalisation-preserving transformation of the system into its set of phase dynamics, and so is automatically a superset of the phase group of any other measurement (all of which can always be coarse-grained into 1 by summing their effects together).

While the existence of an inverse map between sets of effects is sufficient for the equality of their associated phase dynamics, it is not necessary. In particular, if there are already significant (non-phase related) restrictions on the allowed transformations (e.g. as would be the case in a highly asymmetric state-space with no decoherence allowed), loosening the restrictions from the phase dynamics by coarse-graining the set of effects may still not be sufficient to enable further dynamics. In such cases, the associated phase dynamics could be identical, regardless.

Lemma 3.8 also allows us to replace a set of effects associated with the outcomes of a measurement into a form where they correspond to orthogonal vectors in the minimum-dimension representation (§ 2.5). In particular, we can demonstrate that the phase dynamics of a measurement with N effects is identical to the phase dynamics associated with preserving $N - 1$ of those elements and the normalization of the state:

⁹This is always mathematically feasible, even if it is not admitted as an automorphism of the state-space.

Corollary 3.11. *For a measurement N with effects $\{\mathbf{n}_i\}_{i=1\dots n_{\max}}$, the constraints placed on transformations are equivalent to the phase condition on the set of effects $\{\mathbf{m}_i\} = \{\mathbf{n}_i\}_{i=1\dots n_{\max}-1} \cup \{\mathbf{1}\}$, in which we have replaced one of the measurement outcome effects by the unit effect.*

Proof. As N is a measurement, $\mathbf{m}_{n_{\max}} = \sum_i \mathbf{n}_i = \mathbf{1}$. We can thus write a linear transformation between these two sets of effects $\mathbf{m}_j = \sum_i c_{ji} \mathbf{n}_i$, where $c_{ji} = \delta_{ji}$ for $j < n_{\max}$; and when $j = n_{\max}$, then $c_{n_{\max}i} = 1$ for all i . By lemma 3.8, the dynamics associated with this new group S_M are a superset or equal to the dynamics of N , S_N .

However, there also exists the reverse transformation $\mathbf{n}_j = \sum_i \tilde{c}_{ji} \mathbf{m}_i$ where $\tilde{c}_{ji} = \delta_{ij}$ for $j < n_{\max}$ and for the final row $\tilde{c}_{n_{\max}i} = -1$ for $i < n_{\max}$ and $\tilde{c}_{n_{\max}n_{\max}} = 1$ (because $\mathbf{n}_{n_{\max}} = \mathbf{1} - \sum_{i=1}^{n_{\max}-1} \mathbf{n}_i$). Hence both parts of lemma 3.8 are applicable, and so $\{\mathbf{n}_i\}_{i=1\dots n_{\max}}$ and $\{\mathbf{n}_i\}_{i=1\dots n_{\max}-1} \cup \{\mathbf{1}\}$ generate equivalent phase dynamics. $\square$

Corollary 3.12. *The phase dynamics associated with any measurement also preserve the normalization of the state.*

Proof. Follows immediately from the above. $\square$

We conclude this section by writing down one final lemma, which we will find useful for manipulating the phase dynamics of measurements which share effects.

Lemma 3.13. *For a set of effects $\{\mathbf{e}_i\}$ inducing a subset of dynamics $\mathcal{T}$, a linearly dependent effect $\mathbf{f} = \sum c_i \mathbf{e}_i$ can be freely added to the set of effects with no change in $\mathcal{T}$.*

Proof. From lemma 3.8, the set of dynamics induced by $\mathbf{f}$, $\mathcal{T}_F$ is a superset of (or equivalent to) $\mathcal{T}$ such that $\mathcal{T} \subseteq \mathcal{T}_F$. Consider then the union $\{\mathbf{e}_i\} \cup \{\mathbf{f}\}$ and note that from lemma 3.7, the new phase group is given by $\mathcal{T} \cap \mathcal{T}_F$. As $\mathcal{T}_F$ is a superset of $\mathcal{T}$ (i.e. containing all elements in $\mathcal{T}$), then the intersection of $\mathcal{T}_F$ with $\mathcal{T}$ is simply $\mathcal{T}$. Hence, adding a linearly dependent element to a set of effects does not change the allowed phase dynamics of that set. $\square$

3.7 POVMs and general measurements

Let us now apply some of the above lemmata in a more specific context: positive operator-valued measurements (POVMs). In quantum theory, POVM are a more general type of measurement than the usual projective measurement. They are naturally analogous to effects in the GPT framework, as they provide a means of assigning statistics to states without concern for the updated state after the measurement (see Box

2.5 in [8]). In particular, one can construct a POVM with more outcomes than possible for a projective maximal measurement, but such a measurement cannot perfectly and reliably distinguish between more than the maximally allowed number of states (the dimension of the Hilbert space in quantum theory).

Just as the positive elements $\{M_i\}$ in a POVM satisfy $\sum_i M_i = \mathbb{1}$ so that $\text{Tr}(\rho \sum_i M_i) = 1$ for all states ρ , the effects in a GPT measurement sum as $\sum_i e_i = \mathbf{1}$ such that $\sum_i e_i \cdot s = 1$ for all states s . At no point in lemma 3.8 did we make any assumptions about how many (or what sort of) effects were in the measurements (cf. lemma 3.6 which only holds for maximal measurements); and so the results of § 3.6 hold as true for POVMs as they did for projective measurements.

In this section, we will consider how our formalism for phase dynamics applies to non-projective measurements, by way of a couple of examples from quantum theory.

3.7.1 Unambiguous quantum state discrimination

We will first consider a famous example of a non-projective POVM measurement known as *unambiguous quantum state discrimination* [8, 83, 84]. For any two pure quantum states (of the same system) $|\alpha\rangle$ and $|\beta\rangle$, one can construct a POVM of three elements $\{M_\alpha, M_\beta, M_?\}$ that can distinguish between these two states, but not in every attempt of the measurement:

$$M_\alpha = \frac{\mathbb{1} - |\beta\rangle\langle\beta|}{1 + |\langle\alpha|\beta\rangle|}, \quad M_\beta = \frac{\mathbb{1} - |\alpha\rangle\langle\alpha|}{1 + |\langle\alpha|\beta\rangle|}, \quad M_? = \mathbb{1} - M_\alpha - M_\beta. \quad (3.29)$$

M_α will never be observed if the state was prepared in $|\beta\rangle$, and vice versa. However, when $\langle\alpha|\beta\rangle \neq 0$, there is some finite probability that the outcome $M_?$ will be observed, and the measurement will have failed to determine the state either way.

Consider two projective binary measurements, one that distinguishes between $|\alpha\rangle$ and any state $|\bar{\alpha}\rangle$ such that $\langle\alpha|\bar{\alpha}\rangle = 0$, and another between $|\beta\rangle$ and any state $|\bar{\beta}\rangle$ such that $\langle\beta|\bar{\beta}\rangle = 0$, with projectors given respectively by:

$$P_\alpha = |\alpha\rangle\langle\alpha|, \quad P_{\bar{\alpha}} = \mathbb{1} - |\alpha\rangle\langle\alpha|, \quad (3.30)$$

$$P_\beta = |\beta\rangle\langle\beta|, \quad P_{\bar{\beta}} = \mathbb{1} - |\beta\rangle\langle\beta|. \quad (3.31)$$

Using the isomorphism between quantum measurement elements and effects, we can write these projectors in the GPT framework as effects $\{e_\alpha, e_{\bar{\alpha}}\}$ and $\{e_\beta, e_{\bar{\beta}}\}$, and using definition 3.1, we can assign the set of phase dynamics G_α and G_β respectively. These dynamics consist of any allowed operations that do not change the probability of projecting into α (or β), and so will allow more operations than the phase dynamics of

maximal measurements¹⁰ if the dimension of the Hilbert space is greater than two. If the Hilbert space dimension is d , $P_{\bar{\alpha}}$ (and $P_{\bar{\beta}}$) will be a projector of rank $d - 1$, and so the associated phase groups will be of the form $U(1) \oplus U(d - 1)$.

Using the tools provided in this chapter, we can derive a relationship between the phase dynamics associated with the unambiguous quantum state discrimination measurement, and these two related projective measurements.

Theorem 3.14. *The set of phase dynamics admitted by a POVM that performs unambiguous state discrimination between $|\alpha\rangle$ and $|\beta\rangle$ is given by the intersection of phase dynamics associated with the two projective measurements that project onto $|\alpha\rangle$ and onto $|\beta\rangle$.*

Proof. The two projective measurements are associated with the sets of effects $\{e_{\alpha}, e_{\bar{\alpha}}\}$ and $\{e_{\beta}, e_{\bar{\beta}}\}$ respectively. Using corollary 3.11, we can write down sets of effects associated with each of these measurements, equivalent for the purposes of constraining phase dynamics: $\{1, e_{\bar{\alpha}}\}$ and $\{1, e_{\bar{\beta}}\}$ respectively. As the GPT framework does not allow different measurements to normalise differently¹¹, 1 is equivalent in both sets (as may be verified by its action on all s by noting 1 from either measurement must return the same value of $1 \cdot s$). The union of these two sets is $\{1, e_{\bar{\alpha}}, e_{\bar{\beta}}\}$, which from lemma 3.7 has dynamics given by the intersection of the phase dynamics $G_{\alpha} \cap G_{\beta}$.

This set can be related to the set associated with unambiguous discrimination,

$$\{1 - k(e_{\bar{\alpha}} + e_{\bar{\beta}}), ke_{\bar{\alpha}}, ke_{\bar{\beta}}\} \quad (3.32)$$

by the matrix:

$$C = \begin{pmatrix} 1 & -k & -k \\ 0 & k & 0 \\ 0 & 0 & k \end{pmatrix}, \quad (3.33)$$

which admits (for $k \neq 0$) an inverse

$$C^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & \frac{1}{k} & 0 \\ 0 & 0 & \frac{1}{k} \end{pmatrix}. \quad (3.34)$$

Thus, by lemma 3.8 the sets of phase dynamics are equivalent.

We hence see that the phase dynamics of a POVM that performs unambiguous quantum state discrimination is equivalent to the intersection of phase dynamics for

¹⁰ This also illustrates that maximal measurements should not be conflated with projective measurements; although a projective measurement whose matrix elements are all rank one projectors will be maximal.

¹¹ This is effectively another constraint in its own right; but one placed at an even more fundamental level in the framework.

the two projective measurements. □

Corollary 3.15. *For a two-level quantum system, the phase group of an unambiguous quantum state discrimination measurement is trivial unless $\langle \alpha | \beta \rangle = 0$*

Proof. In the Bloch sphere representation, the two non-orthogonal projectors define a fixed plane (whereas a single maximal measurement only defines a fixed axis), and so geometrically, the only allowed operations would be the shrinking or flipping along the single remaining orthogonal axis (in general, given by the map on the density matrix $\Phi : \rho \rightarrow k\rho^T + (1 - k)\rho$). This map necessitates a violation of complete positivity (this may be verified by applying Choi's theorem [41] to Φ), and so the only geometrically valid phase transformations are forbidden by another quantum law. Thus the phase dynamics associated with an unambiguous quantum measurement between two non-orthogonal states on a qubit are trivial. □

3.7.2 Informationally complete measurements

We now examine a second type of non-projective POVM, which can be constructed for all theories in the GPT framework. Consider a system whose state can be completely determined by a set of measurements $\{M_i\}_{i=1\dots N}$, each measurement M_i with the set of effects $\{m_{ij}\}_j$. We can create a new measurement by taking the union of each set of effects, weighted by some non-zero factor (e.g. $\frac{1}{N}$) so that the resultant measurement is correctly normalised.

The statistics built from this compound measurement match those that would be found by randomly choosing a measurement, recording the measurement chosen, and its outcome. Repeated trials of this measurement on identically prepared states would ultimately perform full tomography of the system: every statistic described by the state would eventually be known. For this reason, this type of measurement is known as *informationally complete* (see e.g. [85, 86]).

Lemma 3.16. *The phase dynamics of an informationally complete measurement are trivial.*

Proof. As the measurement contains an effect for every degree of freedom in the state, there are no parts of the state that can change without altering the probability associated with at least one effect. □

Theorem 3.5 (that classical phase dynamics are trivial) could be seen as a corollary of this lemma; for classical systems, a maximal measurement is informationally complete.

3.8 Conclusion

In this chapter, we have seen that it is possible to generalise the concept of phase into the GPT framework by considering the set of transformations that preserve the statistics associated with a particular measurement. This generalised definition reduces to the quantum case when applied to quantum theory, but also allows us to discuss the phase associated with non-quantum theories such as gbits and Spekkens' toy model. Moreover, the concept of phase was shown to be intrinsically related to non-classicality and the existence of complementary measurements. In particular, the phase associated with a maximal measurement helps to identify characteristic behaviours of different theories. We examined how our definition of phase is related to the idea of linear constraints, and derived some relationships between phase groups of similar measurements (such as those sharing outcomes; or those which formed by the coarse-graining of another measurement). We used these relationships to identify the phase dynamics associated with an unusual example of a quantum measurement known as unambiguous state discrimination.

In the following chapters, we will put the mathematical tools developed here to use in the physical context of an interferometer.

Chapter 4

Interference in the GPT framework

The GPT framework has typically been employed to describe abstract degrees of freedom, often imagined as internal degrees of freedom (such as the spin of a particle, or the polarisation of a photon). This picture has proven useful for thinking of generalisations of the EPR experiment (e.g. [54, 65]), addressing questions of information complexity (e.g. [18, 44, 67, 87, 88]), and device-independent cryptography (e.g. [62]).

One could also directly encode spatial degrees of freedom. The choice of a particle to follow one route over an alternative may be described by a bit of information. In quantum theory, which admits coherent superpositions of both possibilities, information about this “which-route?” measurement may be encoded as a qubit. Here, we consider extending such a description to general theories within the GPT framework.

In this short chapter¹, we shall explicitly consider systems that incorporate a spatial degree of freedom throughout: interferometers. We will first revise the quantum behaviour of a Mach–Zehnder interferometer (§ 4.1). Next, using the results of the previous chapter, we will generalise this device into arbitrary theories (§ 4.2). In particular, we will look at couple of non-quantum examples of interference: 3-in 2-out gbits (§ 4.2.1), and Spekkens’ toy model (§ 4.2.2). We shall conclude this chapter with a couple remarks on how we will further abstract the key concept of interferometry in subsequent chapters, and how the treatment here compares with an alternative presentation of interference [89, 90] in the GPT framework (§ 4.3).

4.1 The quantum Mach–Zehnder interferometer

Consider the Mach–Zehnder interferometer being traversed by a single particle (illustrated in fig. 4.1). Let the position of the particle (upper vs. lower branch) be expressed by the $+1$ and -1 outcomes of the Z measurement, such that the quantum state $|0\rangle$ corresponds to the particle being in the upper branch with certainty. The action of the

¹Some of the results in this chapter were published in [1].

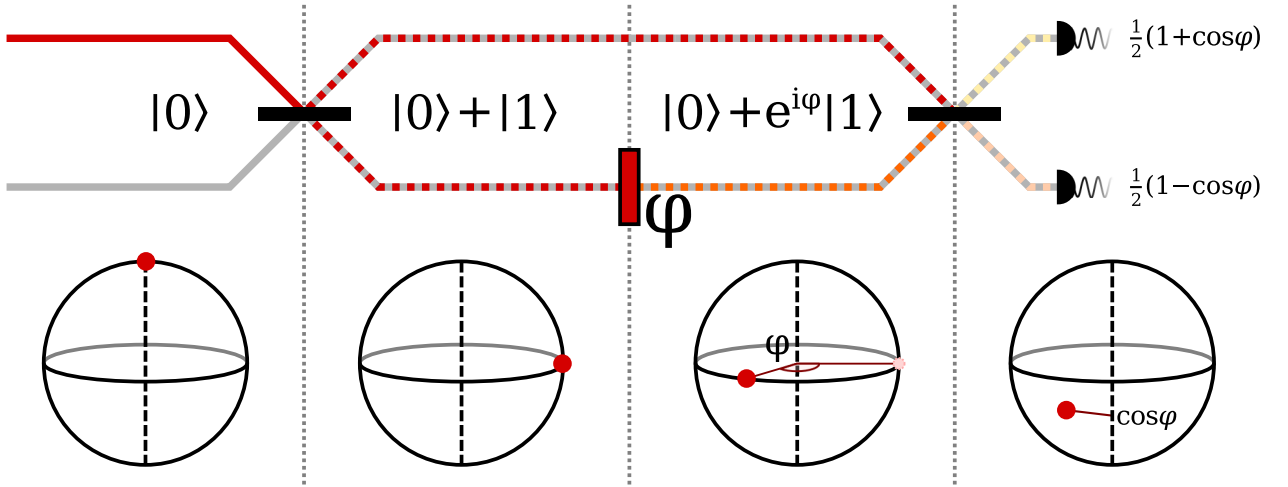


Figure 4.1: A quantum Mach-Zehnder interferometer. The top of the figure is a schematic representation of the experiment. A photon is incident on the first beam-splitter, and enters into a superposition of the two branches. A phase-plate introduced to one of the branches induces a phase that transforms the state. Finally, the two branches recombine with a second beam-splitter, and the photon arrives in one of the two detectors with some probability. The bottom row illustrates the state of the system at each part of the experiment, as a location in the Bloch sphere.

beam-splitter is described by the Hadamard gate

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad (4.1)$$

such that the action of the beam-splitter on a particle incident on the beam-splitter from the upper branch is to place it into a super-position of both outputs $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.

After the beam-splitter, some sample is placed in the path of the lower beam, adding a phase ϕ to the system (e.g. by increasing the optical path length). This constitutes a phase transformation with respect to the ‘which-branch?’ measurement Z , because the operations do not alter the chance of the particle being found in the upper or lower branch at this point. The form² of such an operation on the lower branch may be written

$$U_\phi = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}, \quad (4.2)$$

and it rotates the quantum state around the equator of the Bloch sphere by an angle ϕ .

The particle then enters the final beam-splitter, which performs the inverse operation of the first beam-splitter (which happens to be the same in the quantum case as

²A full justification for this form will be considered in §5.2.

$H^{-1} = H$).

The action of the interferometer as a whole is to create the quantum state $|\text{final}\rangle = HU_\phi H|0\rangle$ from a particle that was initially in the upper branch. We can directly calculate this state to be

$$|\text{final}\rangle = e^{\frac{i\phi}{2}} \begin{pmatrix} \cos\left(\frac{\phi}{2}\right) \\ -i \sin\left(\frac{\phi}{2}\right) \end{pmatrix}. \quad (4.3)$$

From this state, we find that probability of detecting the particle in the upper or lower branch, after it has left the final beam-splitter, is given by

$$P(\text{upper}) = |\langle 0 | \text{final} \rangle|^2 = \frac{1}{2} (1 + \cos \phi), \quad P(\text{lower}) = |\langle 1 | \text{final} \rangle|^2 = \frac{1}{2} (1 - \cos \phi). \quad (4.4)$$

It is reasonable to treat these Z statistics as the *interference pattern* produced by the device, as varying ϕ (such as by extending the length of one branch) will alter them. Furthermore, had the beam-splitters not had a coherent action (and instead mapped the input state into the maximally-mixed classical mixture of the branches, lying at the centre of the Bloch sphere), then the output state would itself be maximally-mixed for all values of ϕ , and so the output statistics would reveal nothing about the choice of ϕ .

Let us now go through the above example again, this time using the GPT formalism. The initial state $|0\rangle$ is written as

$$\mathbf{s}_0 = \begin{pmatrix} \frac{P(X=+1)}{P(X=-1)} \\ \frac{P(Y=+1)}{P(Y=-1)} \\ \frac{P(Z=+1)}{P(Z=-1)} \end{pmatrix} = \begin{pmatrix} \frac{1/2}{1/2} \\ \frac{1/2}{1/2} \\ \frac{1}{0} \end{pmatrix}. \quad (4.5)$$

The two quantum operations H and U_ϕ may be expressed³ as T_H and T_ϕ respec-

³ Because of the constraint that all measurements normalise to the same value, there are in fact unobservable degrees of freedoms available when writing down the transformation, corresponding to the fact that this is a six-dimensional representation of a state with only three degrees of freedom if the state is assumed to be normalised. For example, the top row of the Hadamard gate could have been written as $(k, k, -k, -k, 1, 0)$ where the terms in k would always cancel out when multiplying on a correctly normalised state. As we are only considering the transformations on valid states, these freedoms disappear, and so here we have chosen forms for simplicity: H to minimize the number of non-zero elements, and T_ϕ such that when $\phi = 0$, $T_\phi = \mathbb{1}_6$ explicitly.

tively:

$$T_H = \left(\begin{array}{cc|cc|cc} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{array} \right), \quad (4.6)$$

$$T_\phi = \frac{1}{2} \left(\begin{array}{cc|cc} 1 + \cos \phi & 1 - \cos \phi & -\sin \phi & \sin \phi \\ 1 - \cos \phi & 1 + \cos \phi & \sin \phi & -\sin \phi \\ \hline \sin \phi & -\sin \phi & 1 + \cos \phi & 1 - \cos \phi \\ -\sin \phi & \sin \phi & 1 - \cos \phi & 1 + \cos \phi \end{array} \right) \oplus \mathbb{1}_2. \quad (4.7)$$

T_ϕ is a representation of an element of $SO(2)$ acting on a probability vector. In the expectation value representation, where states are written by Bloch vector $\mathbf{s} = (\langle X \rangle, \langle Y \rangle, \langle Z \rangle)^T$, T_ϕ may be expressed in the more familiar form $\begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \oplus \mathbb{1}_1$. In both representations, it can be seen by inspection that T_ϕ does not alter the statistics associated with Z , as expected for an element of the phase group of Z .

In the probability vector representation, the final state $\mathbf{s}_f$ is written

$$\begin{aligned} \mathbf{s}_f &= T_H T_\phi T_H \mathbf{s}_0 \\ &= \begin{pmatrix} 1/2 \\ 1/2 \\ \hline \frac{1}{2}(1 - \sin \phi) \\ \frac{1}{2}(1 + \sin \phi) \\ \hline \frac{1}{2}(1 + \cos \phi) \\ \frac{1}{2}(1 - \cos \phi) \end{pmatrix}, \end{aligned} \quad (4.8)$$

where the Z statistics are identical to those obtained by the Hilbert space formalism.

4.2 Mach-Zehnder interferometers in the GPT framework

We shall extend the concept of interference to other theories in the GPT framework, by analogy with the network of transformations above.

Theories that can describe a directly analogous Mach-Zehnder interferometer require the following:

- i. There must be a measurement (we shall label Z) corresponding to spatial location (and by corollary, for each branch of the interferometer, there must be an effect which assigns the probability of being in that branch).

- ii. We must be able to prepare at least one state where one outcome of Z is certain.
- iii. There must be a ‘beam-splitter’ transformation T_H , whose effect is to relate the statistics of Z with that of some other measurement.
- iv. There must be a choice of operations $\{T_\phi\}$ in the phase group of Z .

Without item i, it is hard to formulate any concept of “over here vs. over there”; the existence of the effects allows us to represent the detectors at the end of the device. Item ii allows for the initial state of the system to correspond to something in one branch (which we will label as the upper branch), and item iii allows us to prepare the state into a superposition of some kind.

Finally, item iv implies that there is some meaningful operation we can perform, that might affect the output statistics. A Mach-Zehnder interferometer consists of two spatially-disjoint branches, and so between the initial and final beam-splitter, any applied transformations should not alter the probability of being in one branch over the other. If this were not the case, then transformations corresponding to the action on one branch could potentially be used for super-luminal signalling. Suppose an agent has free choice of which transformation to employ on her branch, and this choice could be exercised to change the probability of the particle being found in the remote branch; if another agent is waiting there with measurement device, this can detect if such a manipulation was performed, allowing for instantaneous signalling. For these reasons, we restrict T_ϕ to be a member of the phase dynamics of the position measurement. (Whether any particular phase operation in the phase dynamics of the position measurement can be performed on any particular branch will be discussed in considerable depth in chapter 5. Further relativistic considerations will be elaborated upon in chapter 6.)

Let us, by analogy with the quantum case, restrict our discussion to reversible phase dynamics (i.e. the phase group) for the following examples⁴. In this case, the generalised Mach-Zehnder interferometer produces the following output state:

$$\mathbf{s}_f = T_H^{-1} T_\phi T_H \mathbf{s}_0 \quad (4.9)$$

where T_ϕ is an element of G_Z , the phase group of the position measurement Z (effects $\mathbf{z}_+$ and $\mathbf{z}_-$). The interference fringes are then defined by the probabilities $P(Z = +1) = \mathbf{z}_+ \cdot \mathbf{s}_f$ and $P(Z = -1) = \mathbf{z}_- \cdot \mathbf{s}_f$ respectively. If a theory allows for differing choices of $T_\phi \in G_Z$, and these choices affect the output statistics of Z , then the system is said to exhibit *non-trivial interference*.

⁴Such a set-up could in general admit irreversible dynamics, such as a ‘measure-and-forget’ operation on the two branches, that replaces the coherent state with a mixture (see § 3.4.1)

4.2.1 Example: Interference for 3-in 2-out gbts

Let us consider the above as applied to a gbit with three binary measurements. The T_H in the quantum case (eq. (4.6)) is also a valid transformation on 3-in 2-out gbts, and so we use this for the action of the beam-splitters (noting that $T_H^{-1} = T_H$).

The phase group associated with the Z measurement of a gbit consists of the eight automorphisms of the square:

$$\begin{aligned} T_1 &= \left(\begin{array}{cc|cc} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right) \oplus \mathbb{1}_2, & T_2 &= T_1^2, & T_3 &= T_1^3, & T_4 &= \mathbb{1}_6, \\ T_5 &= \left(\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right) \oplus \mathbb{1}_2, & T_6 &= T_1 T_5, & T_7 &= T_1^2 T_5, & T_8 &= T_1^3 T_5. \end{aligned} \quad (4.10)$$

In this phase group, T_1, T_2, T_3 and T_4 are the proper rotations by some multiple of 90° around the Z -axis; T_5, T_6, T_7 and T_8 are a reflection in the YZ plane followed by a rotation by some multiple of 90° .

Unlike the quantum case, the initial state s_0 is not unambiguously defined (even setting $Z = +1$, still leaves the freedom to choice X and Y statistics), and so we write the output statistics in terms of the input state. The outcome probabilities observable for different choices of T_ϕ are given in table 4.1. As these outputs can be different for different choices of T_ϕ , we say that the gbit MZI can potentially exhibit non-trivial interference patterns. (Whether or not these patterns could be induced by operations that can be meaningfully described as local to either branch in the interferometer will form the main topic of discussion in chapter 5).

Even with the ability to also prepare arbitrary Y statistics for the input state, not every choice of T_ϕ can be distinguished by this particular set-up: T_1 vs. T_6 , T_2 vs. T_5 , T_3 vs. T_8 and T_4 vs. T_7 can not be resolved as they share exactly the same statistics. This is not unique to gbit interferometry. For example, in a quantum MZI, any phase $\phi \in [0^\circ, 180^\circ]$ would be indistinguishable from a phase $\phi' = 360^\circ - \phi$, as even though (for $\phi \neq 0$) these correspond to different elements in $SO(2)$, they produce the same output Z statistics.

On the other hand, a gbit interferometer might allow the ability to prepare two different choices of initial state: s_0 where $P_i(Y = +1) = 1$, and s_0' where $P_i(Y = -1) = 1$. This is in contrast to quantum theory, where there is only one initial state as by choosing the particle to be in the upper branch, the uncertainty restriction mandates that $P_i(Y = +1) = P_i(Y = -1) = \frac{1}{2}$. This means that up to four possible choices of T_ϕ

Phase group element	Upper branch output probability $P_f(Z = +1)$	Lower branch output probability $P_F(Z = -1)$
T_1	$P_i(Y = +1)$	$P_i(Y = -1)$
T_2	$P_i(Z = -1)$	$P_i(Z = +1)$
T_3	$P_i(Y = -1)$	$P_i(Y = +1)$
T_4	$P_i(Z = +1)$	$P_i(Z = -1)$
T_5	$P_i(Z = -1)$	$P_i(Z = +1)$
T_6	$P_i(Y = +1)$	$P_i(Y = -1)$
T_7	$P_i(Z = +1)$	$P_i(Z = -1)$
T_8	$P_i(Y = -1)$	$P_i(Y = +1)$

Table 4.1: Interference observable in a gbit MZI: the output statistics after the final beam-splitter for each phase transformation T_ϕ are written in terms of the initial statistics P_i .

could be distinguished by a gbit MZI (e.g. by varying both Z and Y for the input, one can distinguish all the proper rotations T_1 – T_4), whereas in quantum theory (in which only Z could be varied) elements T_2 and T_4 would give the same (mixed) statistics.

Although an interferometer is typically a device that probes some unknown sample (i.e. choice of phase group element), if it were instead T_ϕ that could be freely and reliably chosen, then the MZI could be used to perform partial tomography on the input state. However, the X statistics do not feature in the interference pattern of the device presented here. This is because of our choice of our beam-splitter transformation T_H , which has a degree of arbitrariness, even in quantum theory (cf. the alternative choices in fig. 2.3). If we had chosen instead a transformation that exchanges the Z and Y statistics (flipping X), a different version of table 4.1 would be derived, corresponding to a slightly different experimental set-up, in which the output statistics depend on the X and Z statistics of the input state. Thus, in this set-up, it is not sufficient to only vary the phase transformations in order to perform state tomography.

4.2.2 Example: Interference in Spekkens' toy model

We can similarly build the statistics for a Mach–Zehnder interferometer described by a single bit in Spekkens' toy model [28, 29] (see § 2.7). There is no operation in this model that exactly implements the quantum version of T_H (one could verify this by exhaustively going over every possible permutation in S_4 and noting that none give the correct change in statistics⁵), but we can pick a closely related T_H that swaps X

⁵Geometrically, the quantum Hadamard is a rotation of 180° about the axis bisecting X and Z ; such a rotation is not a symmetry of the tetrahedron of ontic states in fig. 2.10.

and Z statistics, but does not flip Y :

$$T_H^{\text{Sp}} = \left(T_H^{\text{Sp}}\right)^{-1} = \left(\begin{array}{cc|cc|cc} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{array} \right), \quad (4.11)$$

corresponding to the permutation 1324 that swaps ontic states $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ and $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$.

The phase group of permutations preserving Z measurements are the four permutations (12)(34), leading to a representation of $Z_2 \oplus Z_2$ with elements

$$\begin{aligned} T_{1234} &= \mathbb{1}_6 & T_{2134} &= \left(\begin{array}{cc|cc|cc} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array} \right) \oplus \mathbb{1}_2 \\ T_{1243} &= \left(\begin{array}{cc|cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right) \oplus \mathbb{1}_2 & T_{2143} &= \left(\begin{array}{cc|cc|cc} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array} \right) \oplus \mathbb{1}_2 \end{aligned} \quad (4.12)$$

The output statistics obtained for each of these four phase operations in the MZI circuit $T_H^{\text{Sp}} T_\phi T_H^{\text{Sp}} s_0$ are listed in table 4.2.

Phase group element		Upper branch output probability $P_f(Z = +1)$	Lower branch output probability $P_F(Z = -1)$
T_{1234}	$(\mathbb{1} \oplus \mathbb{1})$	$P_i(Z = +1)$	$P_i(Z = -1)$
T_{2134}	$(X \oplus \mathbb{1})$	$P_i(Y = -1)$	$P_i(Y = +1)$
T_{1243}	$(\mathbb{1} \oplus X)$	$P_i(Y = +1)$	$P_i(Y = -1)$
T_{2143}	$(X \oplus X)$	$P_i(Z = -1)$	$P_i(Z = +1)$

Table 4.2: Interference observable in a Spekkens' toy model MZI: the output statistics of the final beam-splitter in terms of the chosen phase transformation T_ϕ and the statistics P_i of the incident state.

As in the quantum case (and unlike the gbit case), when the system is initially prepared in the upper branch, there is only one allowed state with the appropriate Z statistics, and in this state the outcomes for X and Y measurements are equally likely. Thus, the particular set-up of the interferometer can only tell the difference between the two possibilities T_{1234} and T_{2143} in a single measurement (or with repeated tests between T_{1234} , T_{2143} , and one of T_{1243} or T_{2134}).

The behaviour of a Mach-Zehnder interferometer governed by Spekkens' toy model

will be further discussed in the next chapter (§ 5.7.4), where we shall compare the behaviour with and without the epistemic restriction.

4.3 Discussion of interference in the GPT framework

4.3.1 Further abstraction of the Mach-Zehnder interferometer

We remark that requirement of the existence of a Hadamard gate (iii in § 4.2) is somewhat restrictive, as it puts a very strong symmetry restriction on the state-space. Although it may have been didactically useful for the above examples, in general the expression of interference in terms of measurement following Hadamard gates proves a somewhat cumbersome notation. At the heart of the interferometer is the concept of identifying an physical operation associated with an element of a phase group that has the ability to change the statistics of the state of a spatially-disjoint system. Hence, for the rest of this thesis we will abstract away the initial and final beam-splitters of the MZI (as in § 2.3.1 and fig. 2.4); and instead make the assumptions that:

- i. If a theory admits superposition states, it is possible to prepare them.
- ii. It is possible to measure any statistic associated with the state. (I.e. any effect can be achieved physically.)

The reasonableness of these assumptions may be theory dependent, and could be probed further as a way of ruling out certain physical theories (although one might be challenged to describe as any theory that does not satisfy the above as *operational*.)

However, the main focus of the discussion we will present in chapter 5 and in chapter 6 is primarily concerned with what can (and can not) be done by phase operations, in the region of the device where the branches are separated. Freed from the requirement of explicitly writing the beam-splitters, we shall find it easier to generalise beyond two-branch devices, and to occasionally consider systems where there are an arbitrary number of disjoint paths a traversing particle could take. Because the central element of applying an element of the phase dynamics over a set of branches remains the same, we refer to such devices as n -branch Mach-Zehnder interferometers; we implicitly imagine that there could be some set-up that prepares the necessary superposition states, and that any change in statistics induced by the phase transformation can be measured, and thus constitutes as *interference*.

4.3.2 Other treatments of interference in the GPT framework

An alternative treatment of interference in the GPT framework has been presented by Ududec et al. [89] and Ududec [90], extending the concept of hierarchical families of interference introduced by Sorkin [91, 92].

Conceptually, the interference in their framework may be thought of as a generalization of Young’s Slit experiment in which a coherent light-source sends a particle through a number of apertures, resulting in a diffraction pattern on the other side. In quantum theory, there are no patterns formed by three-slit experiments that can not be described entirely in terms of combinations of the two-slit interference patterns formed by blocking one of the slits⁶. Ududec shows how this construction of ‘filters’ over slits can be expressed in the GPT framework. The largest number of slits whose interference patterns can not be described in terms of combinations of a smaller number of slits is known as the *order* of the interference in that theory; and for general theories this could be higher than two.

This hallmark of non-quantum behaviour was put to further use in [17], where it was shown that the non-existence of interference above second-order (alongside the requirements of symmetry and spectrality) is enough to recover that the system behaves as a Jordan algebra [93, 94]; and when further taken with the dimensional constraint that arises by mandating the existence of an energy measurement that can also be used as a generator of dynamics, one recovers the complex Hilbert space usually associated with quantum mechanics.

This approach to interference is thus somewhat different to that presented in this thesis, where the interference is discussed more in terms of the dynamics of a system. Although constructed to address different questions, our treatments are not incompatible. To bridge the gap, one would have to consider the set of transformations which preserve the “which slit?” measurement as the relevant phase group, alongside some method (e.g. chapter 5) for identifying operations that are local to a particular subset of slits. By examining the group structure that relates the operations local to a set of slits, with the operations local to strict subsets of that set, one should recover the hierarchical structure which forms the core of Ududec’s discussion.

⁶This quantum limit can intuitively be seen as a direct consequence of the Born rule in that evaluating some expression of the form $|\psi_1 + \psi_2 + \psi_3|^2$ only contains pair-wise combinations of $\{\psi_1, \psi_2, \psi_3\}$.

4.4 Conclusion

In this chapter, we saw that there is a natural generalisation of the Mach-Zehnder interferometer into post-quantum theories using the definition of phase provided in chapter 3. We identified that the key aspect of an interference experiment is to induce a measurable change in the output interference pattern by the insertion of an unknown sample into the device. In the GPT framework, the sample can be associated with an element of the phase group associated with the “which branch?” measurement, and the interference pattern is entirely encoded in the output state of the set-up. This observation allowed us to discuss the interference that could be seen in post-quantum theories, including Spekkens’ toy model and gbits, and we saw that these theories admitted different behaviour than the quantum case.

In the next chapter we will further probe the Mach-Zehnder interferometer by considering whether it is possible in general to localise a phase operation to just one branch of the interferometer.

Chapter 5

Interference, locality, and uncertainty

Locality, as a general concept, is concerned with the idea of there being disjoint regions in space. For such a statement to be meaningful, there have to be operational consequences that cause different behaviour depending on which region a system is associated with. The most famously successful theory to concern itself with locality is *special relativity* [30]. A significant restriction that arises from special relativity is ‘no-signalling’, which prohibits super-luminal signalling: operations that transmit information between two points in amount of time faster than it would take light (which itself has a finite speed) to traverse that distance.

Theories within the GPT framework are usually chosen to satisfy this restriction (see §2.6), and as such may be described as ‘non-signalling’. Strictly, ‘no-signalling’ is itself a slight abstraction of the restriction placed by relativity: one never allow choices of operation or measurements on one subsystem to affect any measurable statistic on the other subsystem. In the context of special relativity, this would be equivalent to adding the extra assumption that the events taking place on the subsystems are separated by a space-like interval, and that measurements are made (and the experiments concluded) on quicker time-scale than the time it would take for light to travel from one system to another. This abstraction allows the concept of no-signalling to be used in contexts where relativity (or even space-time in general) is not necessarily clearly defined.

No-signalling is a distinct concept to the Bell-violation ‘non-locality’ (e.g. [52, 54, 57–59, 61, 65]), which was discussed in §2.6.2, which refers more to the surrendering of local realism. There are statistics of P-R boxes that exceed the CHSH-Bell inequality by the maximum amount arithmetically possible, but still do not violate the (relativistic) locality consideration of no-signalling.

In this chapter¹, we shall motivate and introduce another type of physical locality restriction, *branch locality*, which applies to the transformations of spatially disjoint systems such as the branches of a Mach–Zehnder interferometer (§5.1). We will verify that this concept is compatible with quantum theory (§5.2). Then, we will see that branch locality forbids all non-trivial dynamics in a certain type of interferometry set-up (including gbit MZIs) called *fully independent* (§5.3). We will see that by

¹Some of the results in this chapter were published in [2].

placing a certain kind of *conditional restriction* on the state-space, this effect can be mitigated (§ 5.4). We demonstrate that a slight generalisation of the quantum uncertainty principle is one such restriction that allows us to have non-trivial dynamics, and still respect branch locality (§ 5.5). We will elaborate on the relation between conditional restriction and the uncertainty principle using an operational generalisation of mutually-unbiased bases (§ 5.6). In the second half of the chapter, we shall present a geometric way in which to visualise the results, and then use this to probe the behaviour of various non-quantum interferometers (§ 5.7). In particular, we will examine the viability of executing the single-bit Deutsch’s algorithm [21, 22, 26] in non-quantum theories (§ 5.8). Finally, using a couple of examples, we consider how the definition of branch locality might warrant extension (§ 5.9).

5.1 Branch locality

We shall motivate the need for a new locality principle, similar in spirit to non-signalling, by first returning to the example of a Mach-Zehnder interferometer. A traditional wave-optical set-up of the MZI could be used to probe a property of an unknown material by measuring its effect on the optical path length. The interference pattern output from the device allows the measurement of the product of thickness and refractive index of the unknown material. In order to take this reading, the material to be measured is deliberately placed into *just one* of the two branches and the other is left empty as a reference.

In quantum theory, a similar set-up leads to the single-qubit version of Deutsch’s algorithm [21, 22, 26]. In this scenario we imagine ‘black boxes’ placed on *each* of the branches of the interferometer. Each box might have a piece of glass in it, that will apply a phase shift of π to traversing particles. As these phase shifts do not alter the probability of the particle being in one branch rather than the other, their action on the system is in the phase group of the position measurement.

Convention in quantum theory tells us that if we label the state of being in the top branch as $|\text{up}\rangle$ and the lower $|\text{down}\rangle$, then a phase shift in the top branch is given by

$$U_{\text{up}} = e^{i\theta_{\text{up}}}|\text{up}\rangle\langle\text{up}| + |\text{down}\rangle\langle\text{down}|, \quad (5.1)$$

and in the lower branch by

$$U_{\text{down}} = |\text{up}\rangle\langle\text{up}| + e^{i\theta_{\text{down}}}|\text{down}\rangle\langle\text{down}|. \quad (5.2)$$

In this language, the action of the two black boxes corresponds to choosing $\theta_{\text{up}} = \{0, \pi\}$ and $\theta_{\text{down}} = \{0, \pi\}$. Deutsch’s algorithm can determine from the output of a single

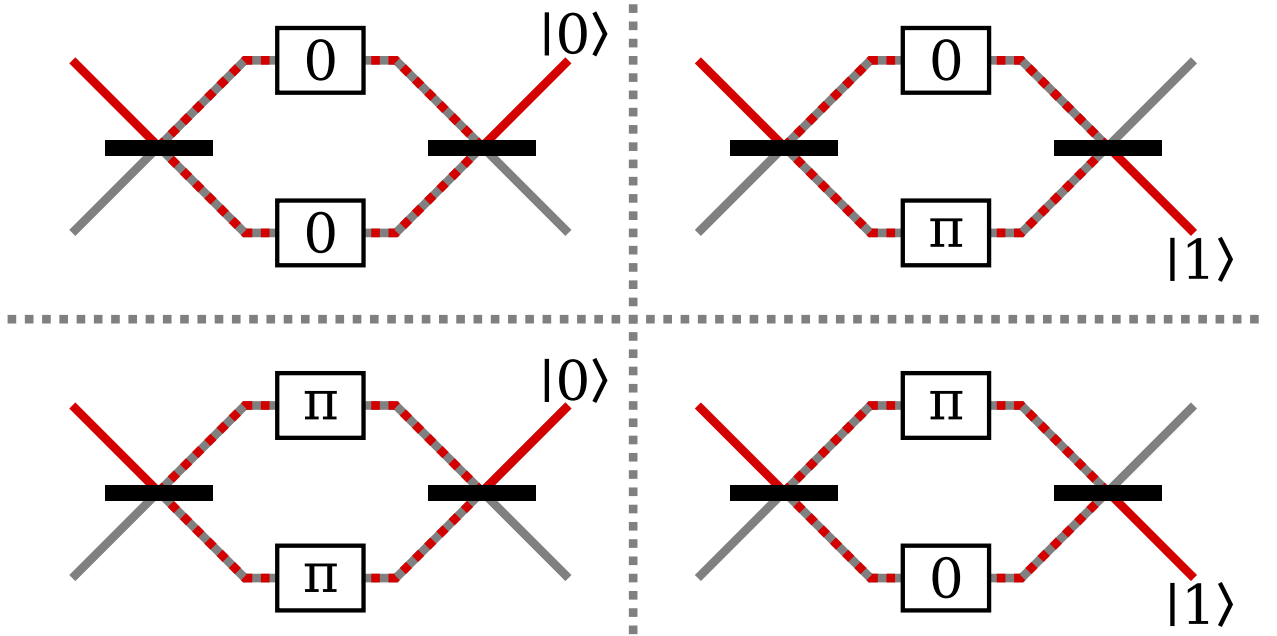


Figure 5.1: Deutsch's algorithm in a Mach-Zehnder interferometer. A single particle, initially in the top branch state $|0\rangle$, passes through a beam-splitter and enters two branches in superposition. Each branch contains a 'black box' phase plate, adding a phase of either 0 or π . The branches are then recombined with a second beam-splitter, and the position measurement is taken as the output. If the phase plates are the same in each branch, then the output is $|0\rangle$; if they differ, then it is $|1\rangle$.

particle passing through the interferometer whether $\theta_{\text{up}} = \theta_{\text{down}}$ or otherwise.

It seems natural that the phase term invoked at a particular branch (up or down) should appear before the ket of that particular branch: as something is being altered about that particular possibility. However, when we step away from the comfortable framework of Hilbert spaces and into the abstracted picture of the GPT framework, it is much less obvious how to assign transformations to have been performed on one branch over another. This is particularly problematic in the single qubit example, where because of the freedom given by the global phase in the unitary transformations, both branches can induce the same full $\text{SO}(2)$ group of operations. Every action on the device can be achieved by actions taken on either branch locally. (In three or more branches, this is no longer the case as will be discussed fully in §5.2).

It is not a priori obvious, whether it should be possible for any GPT theory to apply every transformation via a local action on either branch of a two-level system. Moreover, if not all operations can be done on each branch (such as is the case in three-or-more-level quantum systems), then there needs to be some operational method of determining whether an operation can be performed locally on any particular branch. For this reason, we suggest the following physical condition of locality for branching

interferometers:

Definition 5.1: Branch locality.

If a system has no probability of being detected in a particular branch, then any operations applied on that branch must not affect any statistics of the system. Let e_b be the effect associated with measuring the particle to be in branch b , and state $s_{\neg b}$ be a state that is almost never² in branch b , such that

$$P(Z=b) = e_b \cdot s_{\neg b} = 0. \quad (5.3)$$

All transformations T_b local to branch b satisfy

$$T_b s_{\neg b} = s_{\neg b} \quad (5.4)$$

for all choices of $s_{\neg b}$ with no support on branch b . This condition is known as **branch locality**.

Without such a condition, it would be impossible in practice to ever construct a two-branch (or any finite-branched!) interferometer, because even if the other paths the particle could take were blocked (such that the particle has no chance of traversing them), the actions performed on these paths would still affect the final outcome of the device. The ability to divide space into sections that are part of the experiment and sections are not are why this restriction may be considered as a locality condition: without this condition, one would have to take into account *everything* in order to make predictions about the outcomes of an experiment.

Branch locality does not follow directly from non-signalling. There is a loophole that prevents violation of branch locality automatically violating non-signalling: changes made on a remote system might not be directly measurable on the other branches until all the branches have been brought back together, and so the measurement might still have to be in causal contact (i.e. within the light cone) of whatever operations are done on the branch. However, if branch locality was a trivial consequence of non-signalling, we would not expect to see any new restrictions or phenomena by applying it to the GPT framework, which is already non-signalling by construction. This is not the case, as we shall see in the remainder of this chapter.

²In measure theory, *almost never* means with probability 0, and *almost always* with probability 1. The ‘almost’ concedes that the outcome is conceptually plausible, and might be included in mathematical description (e.g. corresponding to an effect), even if there is no chance of that outcome occurring.

5.2 Branch locality in quantum theory

Before we consider the implications of this restriction in the general case, let us first explicitly explore the effects of branch locality on an interferometer described by quantum theory. The simplest interesting quantum example can be seen in a three-level system (illustrated in Figure 5.2).

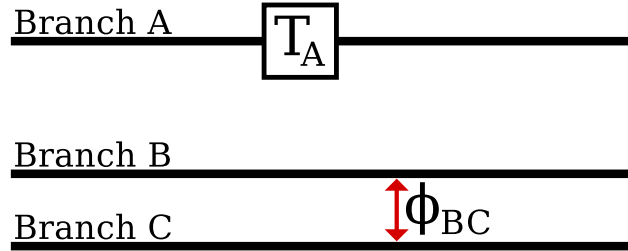


Figure 5.2: Three branch interferometer (initial and final beam-splitters omitted). According to branch locality, operation T_A applied locally on branch A cannot alter the phase ϕ_{BC} between branches B and C.

An operation designated as local to branch A should not alter the phase between branches B and C, as if such an operation could be done, it could induce super-luminal signalling. Consider the scenario in which an agent Alice has branch A, and another agent Bob has B and C. If the state only has support on B and C, then Bob can reliably measure the phase between them. If branch locality was not obeyed, then Alice could induce a phase between the remote branches B and C through an action on her branch A, even though A has no support for the state. Bob could measure this without the need for branch A to ever converge with his other branches.

We shall show below that imposing the restriction of branch locality will rule out such transformations. Although no single branch can implement every transformation, it is true in quantum theory that *any* reversible quantum phase transformation can be implemented by a sequence of transformations, each of which is local to some branch³. This is also proven below.

5.2.1 Reversible quantum transformations and branch locality

For clarity of argument, we shall begin with the discussion of reversible transformations. We can justify our intuition about the form of eq. (5.1) and eq. (5.2) discussed in §5.1.

³It is *not* correct to state that “every operation is local to some branch”. Consider a four-branched interferometer: there is no operation local to a single branch that could change the phase between branches A and B, and also between branches C and D.

Lemma 5.1. *Branch locality implies that the change induced by a reversible quantum operation U_i on a branch i in a set of disjoint branches should be expressed by a complex phase acting only on the ket $|i\rangle$ associated with being in that particular branch:*

$$U_i = e^{i\Phi} \left(e^{i\phi_i} |i\rangle\langle i| + \sum_{j \neq i} |j\rangle\langle j| \right). \quad (5.5)$$

Proof. For a branch i , and a state $\rho_{\neg i}$ that has no support on i such that

$$\text{Tr}(\rho_{\neg i} Z_i) = 0, \quad (5.6)$$

then a branch local operation U_i on branch i always satisfies

$$\text{Tr}(U_i \rho_{\neg i} U_i^\dagger M) = \text{Tr}(\rho_{\neg i} M) \quad (5.7)$$

for any observable M . This implies that

$$U_i \rho_{\neg i} U_i^\dagger = \rho_{\neg i}. \quad (5.8)$$

The most general state that satisfies eq. (5.6) may be written in the Z basis as

$$\rho_{\neg i} = \sum_{\alpha, \beta \neq i} \rho_{\alpha\beta} |\alpha\rangle\langle\beta|. \quad (5.9)$$

According to lemma 3.1, we can express a general quantum phase transformation with respect to Z as $U = \sum_j e^{i\phi_j} |j\rangle\langle j|$, and consider its action on $\rho_{\neg i}$:

$$U_i \rho_{\neg i} U_i^\dagger = \sum_j \sum_{\alpha, \beta \neq i} \sum_k e^{i\phi_j} \rho_{\alpha\beta} e^{-i\phi_k} |j\rangle\langle j| \alpha\rangle\langle\beta| k\rangle\langle k| \quad (5.10)$$

$$= \sum_{\alpha, \beta \neq i} e^{i(\phi_\alpha - \phi_\beta)} \rho_{\alpha\beta} |\alpha\rangle\langle\beta|. \quad (5.11)$$

Therefore, for the condition in eq. (5.8) to be satisfied, we require

$$\sum_{\alpha, \beta \neq i} e^{i(\phi_\alpha - \phi_\beta)} \rho_{\alpha\beta} |\alpha\rangle\langle\beta| = \sum_{\alpha, \beta \neq i} \rho_{\alpha\beta} |\alpha\rangle\langle\beta|. \quad (5.12)$$

For this to hold, it must be true for every element $|\alpha\rangle\langle\beta|$, and this requires $\phi_\alpha = \phi_\beta$ (or for ϕ_α and ϕ_β to differ by an unmeasurable multiple of 2π) for all $\alpha \neq i, \beta \neq i$. A transformation U_i local to branch i thus has the form

$$U_i = e^{i\Phi} \left(e^{i\phi_i} |i\rangle\langle i| + \sum_{j \neq i} |j\rangle\langle j| \right) \quad (5.13)$$

where Φ is some global phase that can be ignored. $\square$

It follows straight-forwardly that all reversible quantum dynamics can be induced by a sequence of local transformations:

Theorem 5.2. *In quantum theory, it is possible to induce every reversible operation in the phase group of a projective measurement by a series of branch local operations independently executed on different branches.*

Proof. Because $\langle i|j\rangle = \delta_{ij}$, it can be seen by inspection that

$$U = \prod_i U_i = \sum_j e^{i\phi_j} |j\rangle\langle j| \quad (5.14)$$

up to some unmeasurable global phase. This is exactly the form of the quantum phase group, given in lemma 3.1. $\square$

5.2.2 Irreversible quantum transformations and branch locality

We shall now probe what the phase and branch locality conditions look like on the most general possible quantum transformations. Interestingly, we find that not every CP-map can be implemented by a series of local actions; it will transpire that there are some types of decoherence which are intrinsically non-local, as their implementation requires co-ordination of some random choice between agents acting on different branches.

Let us begin by extending lemma 3.1 into a restriction on the irreversible phase dynamics of a maximal measurement in quantum theory.

Lemma 5.3. *All Kraus operators associated with a quantum phase transformation of maximal measurement Z are diagonal in the basis associated with that measurement. Conversely, an operation expressed by a set of Kraus operators diagonal in the basis of measurement Z are phase transformations with respect to Z .*

Proof. Consider the most general quantum operation, given by the completely-positive trace preserving map $\rho \rightarrow \sum_a A_a \rho A_a^\dagger$, where $\{A_a\}$ are Kraus operators, that satisfy $\sum_a A_a^\dagger A_a = \mathbb{1}$. The condition for an operation to be a phase transformation of a maximal measurement (described by projectors $\{Z_i\}$) is now written:

$$\text{Tr} \left(Z_i \sum_a A_a \rho A_a^\dagger \right) = \text{Tr} (Z_i \rho) \quad \forall \rho, i. \quad (5.15)$$

After manipulating using the linearity and cyclicity of traces (and the freedom to choose any ρ , including $\rho = \frac{1}{N}\mathbb{1}_N$), we arrive at the restriction

$$\sum_a A_a^\dagger Z_i A_a = Z_i \quad \forall i. \quad (5.16)$$

The terms in this matrix equation can be expressed explicitly in the Z basis as

$$A_a = \sum_{\alpha\beta} (A_a)_{\alpha\beta} |\alpha\rangle\langle\beta|, \quad A_a^\dagger = \sum_{\alpha\beta} (A_a)_{\alpha\beta}^* |\beta\rangle\langle\alpha|, \quad Z_i = |i\rangle\langle i|. \quad (5.17)$$

The left hand side of Eq. 5.16 can thus be written explicitly as

$$\begin{aligned} & \sum_a \sum_{\alpha\beta} \sum_{\gamma\delta} (A_a)_{\alpha\beta}^* (A_a)_{\gamma\delta} |\beta\rangle\langle\alpha| |i\rangle\langle i| |\gamma\rangle\langle\delta| \\ &= \sum_a \sum_{\beta\delta} (A_a)_{i\beta}^* (A_a)_{i\delta} |\beta\rangle\langle\delta|, \end{aligned} \quad (5.18)$$

which we equate to the right-hand side of eq. (5.16), $Z_i = |i\rangle\langle i|$. We may compare element-wise for $|\beta\rangle\langle\delta|$ since $\langle\beta|\delta\rangle = \delta_{ij}$. This yields the conditions:

$$\sum_a (A_a)_{ii}^* (A_a)_{ii} = 1 \quad \forall i, \quad (5.19)$$

$$\sum_a (A_a)_{i\alpha}^* (A_a)_{i\beta} = 0 \quad \forall \alpha \neq i \text{ and } \beta \neq i, \quad \forall i. \quad (5.20)$$

Taking the latter of these conditions, and setting $\alpha = \beta \neq i$, we arrive at $\sum_a (A_a)_{\alpha i}^* (A_a)_{\alpha i} = 0$, which implies that all $(A_a)_{\alpha i} = 0$ when $\alpha \neq i$, because $(A_a)_{\alpha i}$ are complex numbers (and hence every addend in the sum is non-negative). Thus, for phase transformations of the Z measurement, the Kraus operators $\{A_a\}$ are all diagonal in the Z basis, and so may be written in the following form with one fewer index:

$$A_a = \sum_j (A_a)_j |j\rangle\langle j|. \quad (5.21)$$

The converse may be seen by noting that matrices of this form are mutually diagonal with Z_i and hence $[A_a, Z_i] = 0$ for all A_a . Using this, $\sum_a A_a^\dagger A_a = \mathbb{1}$, and the linearity and cyclicity of trace we find:

$$\text{Tr} \left(Z_i \sum_a A_a \rho A_a^\dagger \right) = \text{Tr} \left(\sum_a A_a^\dagger Z_i A_a \rho \right) = \text{Tr} \left(\left(\sum_a A_a^\dagger A_a \right) Z_i \rho \right) = \text{Tr} (Z_i \rho), \quad (5.22)$$

which is exactly the form of the phase condition given in eq. (5.15). $\square$

From this, we can find the additional restriction placed by assuming that a transformation is local to branch i .

Lemma 5.4. *The Kraus operators $\{A_a\}$ associated with a phase transformation of Z that can be applied locally on branch i satisfy:*

$$\sum_a (A_a)_\alpha^* (A_a)_\beta = 1 \quad \forall \alpha \neq i, \beta \neq i. \quad (5.23)$$

Proof. Consider the additional restriction placed by assuming that a transformation of the form in lemma 5.3 is local to branch i :

$$\text{Tr} \left(M \sum_a A_a \rho_{-i} A_a^\dagger \right) = \text{Tr} (M \rho_{-i}) \quad \forall M, \quad (5.24)$$

such that

$$\sum_a A_a \rho_{-i} A_a^\dagger = \rho_{-i}. \quad (5.25)$$

States ρ_{-i} without support on i are still given by form in eq. (5.9). We can use the form of $\{A_a\}$ for phase transformations in eq. (5.21) to explicitly write out the left hand side of eq. (5.25):

$$\begin{aligned} & \sum_a \sum_j \sum_{\alpha, \beta \neq i} \sum_k (A_a)_j^* \rho_{\alpha\beta} (A_a)_k |j\rangle\langle j| \alpha\rangle\langle\beta| k\rangle\langle k| \\ &= \sum_a \sum_{\alpha, \beta \neq i} (A_a)_\alpha^* (A_a)_\beta \rho_{\alpha\beta} |\alpha\rangle\langle\beta|, \end{aligned} \quad (5.26)$$

and so comparing for each $|\alpha\rangle\langle\beta|$ to the right hand side of Eq. 5.25, we obtain the condition

$$\sum_a (A_a)_\alpha^* (A_a)_\beta = 1 \quad \forall \alpha \neq i, \beta \neq i. \quad (5.27)$$

□

When there is only one Kraus operator in the set, eq. (5.23) reduces exactly to the branch locality condition given on reversible quantum transformations in lemma 5.1. Hence, it is reasonable to take this as the irreversible analogue of lemma 5.1 that curtails the ability of any transformation to affect states that are not associated with branch i .

5.2.3 Example branch-local operation: decoherence

One irreversible operation we would like to be achievable through actions performed on a local branch is decoherence. In particular, let us consider an operation that destroys coherence in the relative phase between some branch i and the other branches, through the probabilistic application of a π phase on branch i . A possible set of Kraus operators for this operation is given by:

$$A_i = \sqrt{1-p} \text{diag}(1, \dots, 1, 1, 1, \dots, 1), \quad A_2 = \sqrt{p} \text{diag}(1, \dots, 1, -1, 1, \dots, 1) \quad (5.28)$$

which only differ by element i , and to ensure irreversibility $p \in (0, 1)$.

By inspection, we see this is a phase operation, as it satisfies the condition in eq. (5.21), and moreover A_i satisfies the condition for branch locality on branch i given by eq. (5.23). A physical intuition for this can be seen by considering the action of this operation on a generic density matrix ρ

$$\begin{aligned} \rho &\rightarrow (1-p)A_1\rho A_1^\dagger + pA_2\rho A_2^\dagger \\ \begin{pmatrix} \rho_{11} & \dots & \rho_{1i} & \dots & \rho_{1N} \\ \vdots & \ddots & \vdots & & \vdots \\ \rho_{i1} & \dots & \rho_{ii} & \dots & \rho_{iN} \\ \vdots & & \vdots & \ddots & \vdots \\ \rho_{N1} & \dots & \rho_{Ni} & \dots & \rho_{NN} \end{pmatrix} &\rightarrow \begin{pmatrix} \rho_{11} & \dots & (1-2p)\rho_{1i} & \dots & \rho_{1N} \\ \vdots & \ddots & \vdots & & \vdots \\ (1-2p)\rho_{i1} & \dots & \rho_{ii} & \dots & (1-2p)\rho_{iN} \\ \vdots & & \vdots & \ddots & \vdots \\ \rho_{N1} & \dots & (1-2p)\rho_{Ni} & \dots & \rho_{NN} \end{pmatrix} \end{aligned} \quad (5.29)$$

and noting that the density matrix is unchanged, except for the non-diagonal elements in row i and column i which are replaced with $(1-2p)\rho_{ij}$. This directly demonstrates why this transformation is a type of decoherence: when $p = \frac{1}{2}$, the off-diagonal terms involved with branch i are all set to 0. When a state ρ_{-i} has no support in branch i , all these elements of the density matrix are 0. This allows us to satisfy the definition of branch locality (definition 5.1): because the operation does not change any other parts of the density matrix (which is an operational state), no measurable statistics of the system are changed as a result of the action taken on the unsupported branch i .

By the same reasoning, it can be seen that any set of Kraus operators of the form $\{\sqrt{p_a} \text{diag}(1, \dots, 1, e^{i\phi_a}, 1, \dots, 1)\}_a$ where $\sum_a p_a = 1$ will be a phase operation that satisfies branch locality for a particular branch.

5.2.4 Non-branch local quantum operations

From the above, we have seen that in a quantum interferometer, by combining (possibly probabilistic) branch local operations over the different branches, a very wide range of dynamics are achievable. It would be tempting to conclude that *all* phase operations in quantum theory satisfy branch locality. However, there are still some irreversible dynamics that cannot be implemented by a series of branch local operations, because they implicitly require some sort of communication between agents on different branches.

A counter-example of this can be seen in a four level system by considering the dephasing operation that destroys the coherences between the first two and the last two levels:

$$\begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{21} & \rho_{22} & \rho_{23} & \rho_{24} \\ \rho_{31} & \rho_{32} & \rho_{33} & \rho_{34} \\ \rho_{41} & \rho_{42} & \rho_{43} & \rho_{44} \end{pmatrix} \rightarrow \begin{pmatrix} \rho_{11} & \rho_{12} & 0 & 0 \\ \rho_{21} & \rho_{22} & 0 & 0 \\ 0 & 0 & \rho_{33} & \rho_{34} \\ 0 & 0 & \rho_{43} & \rho_{44} \end{pmatrix}. \quad (5.30)$$

This could be achieved with a coarse-grained measure-and-forget between branches $\{1, 2\}$ and branches $\{3, 4\}$, but to see it more explicitly as phase operations, we consider the following set of Kraus operators:

$$A_1 = \frac{1}{\sqrt{2}} \text{diag}(1, 1, 1, 1), \quad A_2 = \frac{1}{\sqrt{2}} \text{diag}(1, 1, -1, -1). \quad (5.31)$$

Operationally, this procedure can be thought of a single random choice where the outcome of a coin-flip must then be communicated between agents acting on branches 3 and 4 telling them whether to apply a phase gate, ensuring that both agents take the same action. Thus, this is an intrinsically non-local operation, as communication is required between the agents to ensure they act in such a way that does not destroy the coherences between them.

A closely-related operation that does satisfy branch locality would be:

$$A_1 = \frac{1}{\sqrt{2}} \text{diag}(1, 1, 1, 1), \quad A_2 = \frac{1}{\sqrt{2}} \text{diag}(1, 1, -1, 1), \quad (5.32)$$

$$B_1 = \frac{1}{\sqrt{2}} \text{diag}(1, 1, 1, 1), \quad B_2 = \frac{1}{\sqrt{2}} \text{diag}(1, 1, 1, -1). \quad (5.33)$$

where operation A can be done on branch 3 and operation B on branch 4. However, this operation will also set matrix elements ρ_{34} and ρ_{43} to 0, and so it does not achieve exactly the transformation in eq. (5.30).

5.2.5 The Aharonov–Bohm effect

On first inspection, the definition of branch locality may seem to contradict a quantum phenomenon known as the *Aharonov–Bohm effect* [95]. This effect manifests, for example, in a double-slit experiment in which a magnetic solenoid is positioned between the slits. When the solenoid is activated, even though there is no magnetic field outside the solenoid, a measurable change in the interference fringes is observed. Thus, the remote magnet has effected a phase transformation on the system.

This is not in violation with branch locality. The switching on or off of the magnet is an intrinsically global operation. This action changes the magnetic potential⁴ and hence affects the phase on all of the branches, in contrast to the addition of a phase plate which is localised to one particular branch.

Thus, the Aharonov–Bohm effect implements a phase transformation via a global operation (and moreover, will still only affect the measurable statistics between branches which have some probability of containing the particle), whereas branch locality is concerned with operations applied to subsets of branches. Indeed, the same statistics as turning on the solenoid could be simulated by introducing appropriately thick wave-plates at each slit in the device (i.e. adding the same phase locally to each branch). The difference between this sort of *geometric phase* (see e.g. [96]) and other types of phase has been abstracted away by our definition in the GPT framework, which has been set up to consider what the effect of the phase is, rather than where it arose from. To violate branch locality, we would require that a phase plate (calculated to give an equivalent phase as the Aharonov–Bohm effect would, or otherwise from another origin) on a branch with no support still induces a change in statistics.

5.3 Fully independent systems

Having concluded our discussion of branch locality in quantum theory, let us now move on to general theories in the GPT framework. The principle of branch locality may appear at first tautological, but in this section we shall find that it has extreme consequences on a particular type of system: those where knowing the outcome of the position measurement does not place any constraints on any of the other statistics. Such systems we shall call *fully independent* with respect to the position measurement, as will be defined formally in definition 5.2.

⁴For the purpose of calculating the interference pattern, it is the magnetic potential and not the magnetic field strength that has to be known to calculate the dynamics; and hence it is the potential that should be of interest to us in an operational framework.

5.3.1 Example: 2-in 2-out gbit

We shall consider a simple example before moving on to the more general case. A 2-in 2-out gbit (sometimes referred to as a *square bit* because of the shape of its state-space) has two measurements each with two outcomes. Let one of these measurements Z (by convention) be position, and the other measurement X be some “phase”. The outcome statistics of these measurements are free to take any value independently of each other, so long as they normalise to the same value. The state of such a system s can be expressed then in terms of the normalisation $\mathcal{N}$ and the expectation values of the two measurements $\langle Z \rangle$ and $\langle X \rangle$:

$$s = \begin{pmatrix} \mathcal{N} \\ \langle Z \rangle \\ \langle X \rangle \end{pmatrix}. \quad (5.34)$$

Let T be a transformation on such a state, written as the matrix

$$\begin{pmatrix} \mathcal{N}' \\ \langle Z \rangle' \\ \langle X \rangle' \end{pmatrix} = \begin{pmatrix} t_{nn} & t_{nz} & t_{nx} \\ t_{zn} & t_{zz} & t_{zx} \\ t_{xn} & t_{xz} & t_{xx} \end{pmatrix} \begin{pmatrix} \mathcal{N} \\ \langle Z \rangle \\ \langle X \rangle \end{pmatrix}. \quad (5.35)$$

Consider now the phase transformations on such a state. By their very definition (3.1) they must not alter the Z statistics for any state; which means both $\mathcal{N}$ and $\langle Z \rangle$ must not change⁵. Thus, the only freedom remaining in a phase transformation is to change X according to $\langle X \rangle' = t_{xn}\mathcal{N} + t_{xz}\langle Z \rangle + t_{xx}\langle X \rangle$, such that

$$T_{\text{phase}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_{xn} & t_{xz} & t_{xx} \end{pmatrix}. \quad (5.36)$$

We consider now the most general transformation T_{+1} local to the $+1$ branch (the branch that is occupied with certainty when $\langle Z \rangle = +1$). In order to apply the principle of branch locality, we consider how T_{+1} acts on states of the form $(n, -n, \langle X \rangle)^T$ which have no probability of being in the $+1$ branch. Because the X statistics must not change in this case, we see that

$$\langle X \rangle = \mathcal{N}(t_{xn} - t_{xz}) + t_{xx}\langle X \rangle, \quad (5.37)$$

and as this must hold for any value of $\langle X \rangle$, we see that $t_{xn} = t_{xz} = t$ and $t_{xx} = 1$. T_{+1} must also be a valid transformation when we apply it to more general states (i.e.

⁵Recall that the effect associated with normalisation 1 is the ultimate coarse-graining of any measurement; and so from corollary 3.9 normalization is preserved by the phase transformation of any measurement

not restricting $\langle Z \rangle = -n$) although these cases no longer have to obey the restriction in eq. (5.37). T_{+1} transforms the allowed state $(1, 1, 1)^T$ to a state where the new expectation value of X is given $\langle X \rangle' = 2t + 1$, forcing $t \leq 0$ as otherwise $\langle X \rangle'$ exceeds $+1$. Similarly, considering the action on the allowed state $(1, 0, -1)^T$ yields $\langle X \rangle' = t - 1$ which forces $t \geq 0$ as otherwise $\langle X \rangle'$ will be lower than -1 . Taken together, $t = 0$ and hence the only form T_{+1} is allowed to take is

$$T_{+1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (5.38)$$

There are no non-trivial transformations that can be applied locally on the $+1$ branch. The same arguments can be made for the other branch, resulting also in $T_{-1} = \mathbb{1}$. We therefore see that no phase transformations can be implemented locally by 2-in 2-out gbits; square bits make for boring interferometers!

5.3.2 Fully independent systems in general

This total restriction of locally-applicable transformations can be extended to a wider class of systems.

Definition 5.2: Fully independent

A state-space is **fully independent** of a measurement, if fixing the choice of statistics for that measurement only ever reduces the number of degrees of freedom in choice of state by the number of possible outcomes of the measurement.

All gbits are in this category, when the measurement is taken to be one of the canonical set (e.g. along an edge of the hypercube for binary measurements), the statistics of every measurement can be chosen independently of each other. Quantum theory, on the other hand, is not in this category. This is because it contains measurements where knowing the outcome with certainty tells us that the system is in a particular pure state (e.g. for a qubit, knowing that $P(Z = +1) = 1$ tells us that the system is in state $|0\rangle$), and so all remaining degrees of freedom in the choice of state have been removed.

We can now make the general statement:

Theorem 5.5 (The dynamics of fully independent systems cannot be localised.). *An interferometer described by a state-space that is fully independent with respect to the*

position measurement has no non-trivial⁶ dynamics that can be localised to any (strict) subset of branches.

Proof. The proof follows from noting that the restriction placed by branch locality (eq. (5.4)) has the form of an eigenvector equation with eigenvalue +1, and the freedom in choice of state guaranteed in fully independent theories allows us to find a large set of independent vectors to which this restriction can be applied. One then finds a number of independent +1 eigenvectors of a matrix equal to the dimension of the transformation matrix, which mandates that this matrix is identity.

Consider a fully independent state-space with a position measurement Z taking one of N possible values, and a set of other measurements $\{X_i\}$. In the GPT framework each measurement is normalised to the same value, and so each X_i measurement gives one fewer degree of freedom than its number of outcomes. In total there are $M = \sum_i (\text{outcomes}(X_i) - 1)$ degrees of freedom associated with the X measurements, and thus $D = N + M$ degrees of freedom overall. It is helpful for this argument to adopt the minimum-dimension notation introduced in §2.5, and write the most general state $\mathbf{s}$ as:

$$\mathbf{s} = \begin{pmatrix} \frac{\mathcal{N}}{P(Z = z_1)} \\ \vdots \\ \frac{P(Z = z_{N-1})}{P(X_1 = x_1)} \\ \vdots \end{pmatrix}, \quad (5.39)$$

where there are M components in the X measurement portion of the vector, and N in the Z measurement portion (incorporating the normalisation $\mathcal{N}$). The most general transformation can be written as a $D \times D$ square matrix acting on such a vector.

Take a subset of $1 \leq K < N$ branches, $\mathcal{K}$. We associate the effect $\mathbf{e}_{\mathcal{K}}$ with the state being in this subset (and effect $\mathbf{e}_{-\mathcal{K}}$ with branches outside this subset, such that $\mathbf{e}_{\mathcal{K}} + \mathbf{e}_{-\mathcal{K}} = \mathbf{1}$). Let $T_{\mathcal{K}}$ be a phase transformation applied within $\mathcal{K}$.

Consider when $T_{\mathcal{K}}$ acts on a state $\mathbf{s}_i$ with support only on branch $i \notin \mathcal{K}$ (explicitly $\mathbf{e}_i \cdot \mathbf{s}_i = \mathbf{1} \cdot \mathbf{s}_i = \mathcal{N}$ and $\mathbf{e}_{\mathcal{K}} \cdot \mathbf{s}_i = 0$). Applying the restriction of branch locality to this state allows us to generate the equation $T_{\mathcal{K}} \mathbf{s}_i = \mathbf{s}_i$, which is an eigenvector equation with eigenvalue +1. Because the state-space is fully independent, we can find $M+1$ linearly independent choices of $\mathbf{s}_i$ satisfying the above: one for each of the M freedoms in X statistics, and an additional one from the normalisation. (This can be best seen in the

⁶Non-trivial is used in the sense “other than the identity” (definition 2.8).

minimal representation discussed in §2.5. If the statistics are free to take any value in the range $[0, 1]$, such as is allowed for qbits, these vectors for branch $i = 1$ would be,

$$\begin{pmatrix} \frac{1}{1} \\ \frac{1}{1} \\ 0 \\ \vdots \\ 0 \\ \frac{1}{1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \dots \begin{pmatrix} \frac{1}{1} \\ \frac{1}{1} \\ 0 \\ \vdots \\ 0 \\ \frac{1}{0} \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} \frac{1}{1} \\ \frac{1}{1} \\ 0 \\ \vdots \\ 0 \\ \frac{1}{0} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (5.40)$$

The final vector is linearly independent of the other M , and must be included⁷ to fully span the X statistics for arbitrary normalisations.)

We have spanned the entirety of X statistics within one branch. We can now turn our attention to the remaining $N - K - 1$ branches, that are neither branch i nor in the set $\mathcal{K}$. Each of these contributes one more vector independent to those already found, all of which are restricted by branch locality. This leads to a total of $M + N - K$ independent vectors restricted by branch locality.

There are two ways to account for the final K eigenvectors. The first is to note that, as a topological consequence of state-spaces being described by compact convex sets, every valid linear GPT transformation acting endomorphically on such a set must have at least one stationary state, following from the Brouwer fixed point theorem (see e.g. [97, 98]⁸). Moreover, as a phase transformation does not change the Z statistics, one could construct another convex set from the original state-space by discarding the Z measurement, and a new (affine) transformation $T_{\bar{Z}}$ based on the action of T for a fixed choice of Z statistics. This transformation will also have a fixed point from Brouwer's theorem, and it will also be a fixed point of T on the full state-space for that particular choice of Z statistics. Hence one can find a stationary state with support on each of the K branches, and associate these with the final eigenvalues of the transformation T .

The second way is to explicitly show that any state with support on a branch in $\mathcal{K}$ will be ultimately taken outside the valid state-space unless it is also not changed by

⁷ As an illustrative counter-example, using linear combination of eq. (5.40) attempt to form the valid state that is in branch 1 with certainty, and also has both $P(X_1 = 1) = 1$ and $P(X_2 = 1) = 1$, but do not include the final vector of where all X elements are 0. Every attempt with the correct X statistics will necessarily have a normalisation greater than 1.

⁸Cf. the Perron–Frobenius theorem applied to stochastic matrices [32, 99] with similar consequences.

$T_{\mathcal{K}}$. Consider a fully normalised set of independent vectors, $\mathbf{s}_1 \dots \mathbf{s}_K$ each associated with support on a different branch in $\mathcal{K}$. We write the vectors we have found outside of $\mathcal{K}$ as $\mathbf{s}_{K+1} \dots \mathbf{s}_D$, and remember that these are all invariant under $T_{\mathcal{K}}$. Together these form a set of D independent vectors, and any state may be expressed as a linear combination of them:

$$\mathbf{s} = \sum_{i=1}^{i=K} c_i \mathbf{s}_i + \sum_{j=K+1}^{j=D} c_j \mathbf{s}_j, \quad (5.41)$$

for real numbers $c_1 \dots c_d$.

Consider then the action of $T_{\mathcal{K}}$ on some state $\mathbf{s}_{\alpha}$ where $\alpha \in \mathcal{K}$:

$$T_{\mathcal{K}} : \mathbf{s}_{\alpha} \rightarrow T\mathbf{s}_{\alpha} = \sum_{i=1}^{i=k} c_i^{(\alpha)} \mathbf{s}_i + \sum_{j=k+1}^{j=d} c_j^{(\alpha)} \mathbf{s}_j, \quad (5.42)$$

We invoke the phase restriction, that Z statistics are preserved by $T_{\mathcal{K}}$, and so note that the inner product with any effects associated with the Z measurement must be conserved. For effect $\mathbf{e}_{\alpha}$ associated with measuring support on branch α , we note that $\mathbf{e} \cdot \mathbf{s}_{\alpha} = 1$, and for all other vectors $\mathbf{e} \cdot \mathbf{s}_i = 0$ when $i \neq \alpha$. Applying this effect to both sides of the transformation in Equation 5.42, we find $1 = c_{\alpha}^{(\alpha)}$. Furthermore, by taking effects $\{\mathbf{e}_{\beta}\}$ associated with the remaining branches in $\mathcal{K}$, we find that $0 = c_{\beta}^{(\alpha)}$ when $\alpha \neq \beta$. We can thus rewrite the action of $T_{\mathcal{K}}$ on some state $\mathbf{s}_{\alpha}$ as:

$$T_{\mathcal{K}} : \mathbf{s}_{\alpha} \rightarrow T\mathbf{s}_{\alpha} = \mathbf{s}_{\alpha} + \mathbf{s}_{\text{rest}}^{(\alpha)}, \quad (5.43)$$

where $\mathbf{s}_{\text{rest}}^{(\alpha)} = \sum_{j=k+1}^{j=d} c_j^{(\alpha)} \mathbf{s}_j$.

Let us now consider applying $T_{\mathcal{K}}$ to this output state. $\mathbf{s}_{\text{rest}}^{(\alpha)}$ is composed of linear combinations of $+1$ eigenvectors of $T_{\mathcal{K}}$ and so $T_{\mathcal{K}} \mathbf{s}_{\text{rest}}^{(\alpha)} = \mathbf{s}_{\text{rest}}^{(\alpha)}$. Hence, after n applications of $T_{\mathcal{K}}$, state $\mathbf{s}_{\alpha}$ is transformed to state $\mathbf{s}_{\alpha} + n \mathbf{s}_{\text{rest}}^{(\alpha)}$. For $T_{\mathcal{K}}$ to be a valid transformation, it must not take a valid state to a state outside the allowed state-space (i.e. must not take valid state $(T_{\mathcal{K}})^{n-1} \mathbf{s}_{\alpha}$ to invalid state $(T_{\mathcal{K}})^n \mathbf{s}_{\alpha}$). This will only hold for all values of n if $\mathbf{s}_{\text{rest}} = \mathbf{0}$, as otherwise, there will eventually be a probability above the maximum allowed within the theory, or below the minimum. Hence, the effect of $T_{\mathcal{K}}$ on $\mathbf{s}_{\alpha}$ is $T_{\mathcal{K}} : \mathbf{s}_{\alpha} \rightarrow \mathbf{s}_{\alpha}$ and we have found another $+1$ eigenvector associated with branch α . This can be repeated for each of the K branches in $\mathcal{K}$ to generate K more $+1$ eigenvectors associated with the phase restriction. (These eigenvectors are necessarily independent of those we have already found, as they have support on different branches).

Because we have now found $D = N + M$ independent vectors $\{\mathbf{s}_i\}$, we can express any state in our theory as a linear combination of such vectors $\mathbf{s} = \sum_i a_i \mathbf{s}_i$. The effect, however, of transformation $T_{\mathcal{K}}$ is to take $T : \mathbf{s} \sum_i a_i T\mathbf{s}_i \rightarrow \sum_i a_i \mathbf{s}_i = \mathbf{s}$, and hence

there are no states that can be changed by any operation performed on the subset of branches $\mathcal{K}$. This means that the only transformation $T_{\mathcal{K}}$ consistent with branch locality is the identity.

This argument has been made for the most general strict subset $\mathcal{K}$. Hence for systems with state-spaces fully independent of the position measurement, absolutely no non-trivial dynamics can be localised. $\square$

The above proof even works when $K = N - 1$ (i.e. attempting to determine the operations that can be said to *not* involve one of the branches), as we only had to consider a single branch of the interferometer where there was no support for the particle to span the entire set of X statistics, and hence rule out non-trivial transformations. The proof does not work if we take $\mathcal{K}$ to be the total set of branches (no longer strictly a subset), as this is the only situation where there is nowhere in the system in which branch locality can be applied. In this case, we can only generate N restrictions, associated with the preservation of Z statistics, and so the capacity for meaningful dynamics is recovered only in the most global scenario.

Thus, we have shown that branch locality, when applied to fully independent set-ups, places extreme restrictions on dynamics that can be localised to any particular part of the system, making the viability of meaningful gbit interference very unlikely in a Mach-Zehnder interferometer.

5.4 Fully conditionally-restricted systems

Theorem 5.5 stated that for gbts, by imposing branch locality we completely curtail the possible phase dynamics. On the other hand, we showed in § 5.2 that in quantum theory, all reversible and a wide range of irreversible phase transformations can be performed without violating branch locality. Under the same locality constraints, these two different theories have vastly different levels of dynamics admissible in an interferometer, and we now probe if there is an operational explanation for this.

There was one a stage in the proof of theorem 5.5 that holds true for gbts, but does not apply to quantum theory: the ability to select many independent state vectors associated with the particle being in the same branch of the interferometer. As the ability to find more than one independent vectors in the form of eq. (5.40) is not allowed in quantum theory, significantly less constraints are placed on the dynamics as a consequence of branch locality.

We shall abstract this property into general theories. First of all, we define the opposite of *fully independent* (cf. definition 5.2):

Definition 5.3: Fully conditionally-restricted

A state-space is **fully conditionally-restricted** by a particular measurement if fixing the measurement such that any outcome occurs with certainty is sufficient to completely specify the state.

This property is termed *conditional* because whether the restriction applies depends conditionally on the statistics of one of the measurements. This restriction could also be considered as ‘partial self-duality’, in that the effects associated with the position measurement outcomes should only correspond to one state.

In a similar manner as for the fully independent case, we can make a statement about the localisability of phase dynamics (that is, the ability to perform phase transformations with access to only a subset of branches) for systems governed by fully conditionally-restricted theories.

Theorem 5.6. *The phase dynamics of an interferometer where the state-space is fully conditionally-restricted by the position measurement can be localised to a strict subset of branches.*

Proof. Consider a phase transformation $\mathcal{T}$, and let $\{\mathbf{e}_i\}_{i=1\dots N}$ be the effects associated with the position measurement between the N branches. Recall that for $\mathcal{T}$ to be a phase transformation, it must satisfy N conditions of the form:

$$\mathbf{e}_i \cdot \mathbf{s} = \mathbf{e}_i \cdot \mathcal{T}\mathbf{s} \quad \forall \mathbf{s}. \quad (5.44)$$

Consider the transformation localised to a strict subset of branches $\mathcal{K}$ that contains all but one of the N branches, i . Writing the set of states that have no support on any branches in $\mathcal{K}$ as $\{\mathbf{s}_{\neg\mathcal{K}}\}$, we restate the branch locality condition:

$$\mathbf{s}_{\neg\mathcal{K}} = \mathcal{T}\mathbf{s}_{\neg\mathcal{K}} \quad (5.45)$$

The set of states with no support on $\mathcal{K}$ is exactly the set of states with support on branch i . Because the theory is fully conditionally-restricted, specifying that the state is in branch i with certainty is sufficient to completely determine the state, and so there is one unique state $\mathbf{s}_i$ in this set.

This one state restricted by branch locality puts no further restrictions on the eigenvalues of $\mathcal{T}$ than those already present from the phase conditions in eq. (5.44). As discussed in theorem 5.5, we can find a stationary state (i.e. a +1 eigenvector of $\mathcal{T}$) associated with any configuration of position statistics associated with the phase group, and we can do this for a state localised to each of the N branches. Moreover, as there is only one choice of state localised to that branch in a fully conditionally-restricted theory, these stationary states are exactly those we have already already

accounted for in eq. (5.44).

Thus, there are *no* additional constraints placed by branch locality on the phase transformation $\mathcal{T}$, and hence any transformation in the phase dynamics can be localised to a strict subset of branches. $\square$

This result is particularly strong when applied to two-level systems (such as the standard Mach–Zehnder interferometer).

Corollary 5.7. *For a two-level system described by a theory that is fully conditionally-restricted by the position measurement, all phase dynamics can be induced by transformations that are local to just one of the branches only, and this is true for both branches.*

Proof. For a two-level system, the two largest strict subsets are the two sets that contain only one branch. The argument of the above theorem can then be made for each of these sets, hence allowing all of the phase dynamics of a fully conditionally-restricted theorem to be implementable on a single branch of the interferometer. $\square$

5.5 Quantum-like uncertainty relations

One way a state-space can be made fully conditionally restricted by a measurement is if that measurement is included in an informationally complete set⁹ which obeys a generalised uncertainty relationship (cf. appendix A).

Definition 5.4: **Quantum-like uncertainty relation**

A set of measurements obeys a quantum-like uncertainty relation if for every state where one of these measurements has an outcome that occurs with certainty, the outcomes of all other measurements are given by a uniform random distribution.

Lemma 5.8. *If a theory imposes a quantum-like uncertainty relationship on an informationally complete set of measurements, then the state-space is fully conditionally restricted by every measurement in that set.*

Proof. When one of the measurements takes a known value, all the other measurements are constricted to the uniform distributions, and so there is only one state associated with this particular outcome of the measurement. As the set of measurements is informationally complete, this totally specifies the state and hence the theory is fully conditionally-restricted with respect to every measurement in the set. $\square$

⁹I.e. a set of measurements that allow for complete tomography of any state in the theory.

A quantum-like uncertainty relationship, whilst sufficient, is not a *necessary* condition for a theory to be fully conditionally-restricted. If one over-defines the state with a set of measurements (including Z , the measurement which the others are to be conditionally-restricted by), there could still be some measurements that take non-uniform fixed statistics when Z has a definite outcome. For example, consider describing a qubit by four measurements: the usual X , Y , Z and the additional W which reflects a measurement on the Bloch sphere bisecting X and Z . Consider the state $|0\rangle$. For such a state, $\langle Z \rangle = +1$ and $\langle X \rangle = \langle Y \rangle = 0$. However, $\langle W \rangle$ will not be zero, and hence is not uniformly random. Despite this, it is still constrained to only take one set of statistics; there are no other states which share the same X , Y and Z statistics (or crucially, even just the same Z statistics) that are allowed differing values of W . Hence, even when over-determined and not satisfying a quantum-like uncertainty relationship, the theory is still fully conditionally restricted.

The above example also motivates why being fully conditionally-restricted should be thought of as a property of a system in the theory, whereas the quantum-like uncertainty relation is a property of a (sub)set of measurements within that theory.

Indeed, for quantum systems with a number of levels that is not prime (or a power of a prime), the maximum number of mutually unbiased bases is not known [100], and hence forming an informationally complete set of measurements that obey a quantum-like uncertainty relation is not a trivial problem. The general classification of quantum theory as a fully conditionally-restricted theory would be tricky if it relied on the states always being expressible in a particular set of bases. Fortunately, this is not the case, as we can see in this generalisation of the above example:

Lemma 5.9. *All systems described by quantum theory are fully conditionally restricted by any maximal measurement on that system (regardless of the other measurements in the informationally complete set).*

Proof. The proof is a straight-forward consequence of the self-duality in quantum physics between pure states and rank-one projectors. Any maximal measurement outcome may be associated with just one single pure state, which fixes all the other statistics.

Thus, if an outcome of a maximal measurement is associated with some pure state $|\psi\rangle$, any other outcomes probabilities from some POVM (M) have probabilities given by $P(m|\psi) = \text{Tr}(M|\psi\rangle\langle\psi|)$, and no other degrees of freedom are admitted. Because one maximal measurement outcome is sufficient to find a unique $|\psi\rangle$, and $|\psi\rangle$ encodes the entire statistics of the state, by fixing the outcome of one maximal measurement the entire state is fixed, and the theory is hence fully conditionally restricted, regardless of the other measurements. $\square$

5.6 Mutually unbiased measurements

In quantum theory, the uncertainty principle and the concept of mutually unbiased bases are intrinsically linked. For a Hilbert space of dimension d , two bases $\{|x_i\rangle\}$ and $\{|y_j\rangle\}$ (associated with X and Y measurement respectively) are said to be mutually unbiased if

$$|\langle x_i | y_j \rangle|^2 = \frac{1}{d} \quad \forall i, j. \quad (5.46)$$

One sees that for a state prepared as $|x_i\rangle$, corresponding to the only state where outcome i of the X measurement is certain, making a Y measurement could return any outcome with equal probability. This is true for all i . Likewise, choosing a state $|y_j\rangle$ yields the same behaviour for X measurements. A set of mutually unbiased bases is thus a collection of measurements, which pairwise satisfy this behaviour. We thus see that quantum definition of mutually unbiased bases leads directly to a quantum-like uncertainty relation (definition 5.4).

For GPTs in general, an uncertainty relationship might not hold (such as for gbit state-spaces), so it is not obvious if there should be an immediate generalisation of mutually unbiased bases.

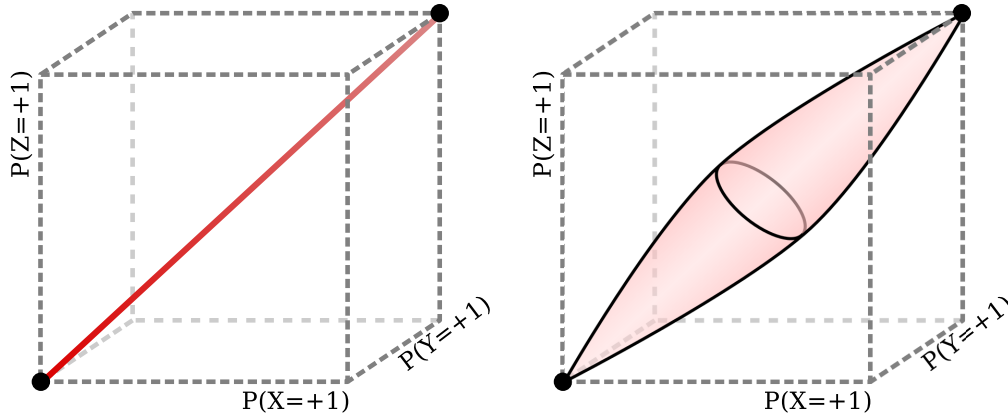


Figure 5.3: Example non-mutually unbiased state-spaces. LHS illustrates a state-space where $\langle X \rangle = \langle Y \rangle = \langle Z \rangle$, and making any measurement is sufficient to totally determine the state. Such a system is effectively classical. RHS illustrates a similar state-space, but where there is some freedom in choice of all three measurement statistics.

A gbit may be described by a set of measurements where knowing the outcome of one measurement will not reveal any information about the outcome of the other measurements in the set if originally the state could have been any allowed state with equal likeliness. In contrast to this, take the example on the LHS of fig. 5.3, for a three-measurement theory where the state-space is constricted to $\langle X \rangle = \langle Y \rangle = \langle Z \rangle$. Here, it is manifestly obvious that making just one of the measurements is sufficient

to determine the entire state. Due to the strict constraints on the states allowed, representing states in this theory using three measurements is redundant, and one could represent the state of a system with these freedoms with a single classical bit. It would be fair to describe the three measurements in this system as totally biased.

Measurements can be ‘biased’ in a way that is not as clear cut as perfect correlation. Consider the similar example on the RHS of fig. 5.3 with a spindle-shaped state-space, such that the statistics of the measurements displayed are still highly correlated, but now there is some freedom in choice of state even after fixing one of the measurements.

Contrast this with the three Pauli measurements on a qubit. Without any prior knowledge of the state, knowing the statistics of Z reveals nothing about what outcome an X or Y measurement will give. Even when we know $Z = +1$, thus fixing the X and Y statistics, these statistics are fixed in such a way that either outcome $+1$ or -1 is still equally probable. On the other hand, for the examples in fig. 5.3 measuring $Z = +1$ would ensure that $X = Y = +1$ also.

Such a bias (or lack thereof) is a property of the measurements used to express the state, rather than of the theory itself. Suppose for the spindle state-space, we were to pick new measurements (ignoring normalisation) $Z' = X + Y + Z$ which measures along the diagonal, and then two extra measurements ($X' = Y - Z$ and $Y' = X - \frac{1}{2}(Y + Z)$), which are orthogonal to the diagonal in the diagram. The state-space, expressed in this new basis, will no longer be slanted, but rather instead look like a regular quantum state-space with a limit on the maximum coherence in a superpositions.

We now suggest a possible definition for a set of measurements to be deemed ‘unbiased’ within the GPT framework.

Definition 5.5: Mutually unbiased measurements

A set of measurements are **mutually unbiased** if for all allowed states, the outcome probabilities of one measurement in the set can be permuted to form another valid state without having to change the statistics of any other measurement in the set.

We can see that this definition implies the quantum definition (eq. (5.46)), using a proof similar in spirit to that of lemma 5.9, drawing on the property of quantum theory that knowing the outcome of a maximal measurement is sufficient to completely determine a quantum state.

Lemma 5.10. *For quantum theory, the above GPT definition for mutually unbiased measurements reduces to the quantum definition of mutually unbiased bases.*

Proof. Consider a measurement X with outcome x_j associated with eigenstate $|x_j\rangle$. The statistics for another measurement Y with outcomes $\{y_i\}$ and eigenstates $\{|y_i\rangle\}$,

when made on the X eigenstate are uniquely given by

$$P(Y = y_i | X = x_j) = |\langle x_j | y_i \rangle|^2. \quad (5.47)$$

If we now consider the Y statistics for another X eigenstate $|x_k\rangle$, there will be a new set of Y statistics given by

$$P(Y = y_i | X = x_k) = |\langle x_k | y_i \rangle|^2, \quad (5.48)$$

and these will be the only Y statistics we are allowed to associate with a state corresponding to X measurement outcome x_k .

Suppose that X and Y are mutually unbiased, as defined by definition 5.5. Then, we should have freedom to permute x_j with x_k without being forced to update any of the other statistics. Hence, there must be a state where $X = x_k$, but the Y statistics for any outcome y_i is given by $|\langle x_j | y_i \rangle|^2$, and because there is only one set of statistics associated with each X outcome, we thus conclude that

$$|\langle x_1 | y_i \rangle|^2 = |\langle x_2 | y_i \rangle|^2 = \dots = |\langle x_d | y_i \rangle|^2 \quad \forall y_i. \quad (5.49)$$

We can make exactly the same argument for pure states of the Y measurement such that

$$|\langle x_j | y_1 \rangle|^2 = |\langle x_j | y_2 \rangle|^2 = \dots = |\langle x_j | y_d \rangle|^2 \quad \forall x_j, \quad (5.50)$$

and hence find that the square of the inner product for every pair $|x_j\rangle, |y_i\rangle$ is the same.

For normalised states over a Hilbert space of dimension d we see that

$$\sum_{i=1}^d |\langle x_j | y_i \rangle|^2 = 1 \quad \forall |x_j\rangle \quad (5.51)$$

and we can replace the different terms in the sum with any one term repeated d times, to find that $d|\langle x_j | y_i \rangle|^2 = 1$ for all x_j and y_i . Thus, we arrive at the quantum statement for two bases to be mutually unbiased:

$$|\langle y_i | x_j \rangle|^2 = \frac{1}{d} \quad \forall x_j, y_i \quad (5.52)$$

□

This restriction imposes freedom to permute any measurement in the set independently, and hence requires a high level of symmetry in the state-space. For this reason, the measurements that form the asymmetrical state-spaces in fig. 5.3 are ruled out. In both cases, the state $(\langle X \rangle, \langle Y \rangle, \langle Z \rangle)^T = (1, 1, 1)^T$ is admitted, but the state $(\langle X \rangle, -\langle Y \rangle, \langle Z \rangle)^T = (1, -1, 1)^T$ is not.

It should be noted that this restriction is on the whole state-space as described by a set of measurements rather than any individual states within it. For example, a square gbit (with X and Z binary measurements) admits corner states such as $(+1, +1)^T$, where the outcome of either measurements could be perfectly predicted. However, this theory also admits the states where one of the outcomes is swapped: $(+1, -1)^T$ and $(-1, +1)^T$, and hence X and Z still satisfy the condition of definition 5.5 to be mutually unbiased.

Finally, we note that the definition does not require the permutation to be physically implementable by an allowed transformation of the theory. Although we require the permutation of any measurement statistics to map valid states to valid states, it is not a problem if this map were to be ruled out by some other imposed restrictions. For example, in a qubit where Pauli operators X , Y and Z form a set of mutually unbiased measurements, replacing Y with $-Y$ takes valid states to valid states, satisfying the definition, but is not an allowed quantum operation because it violates complete positivity.

We can relate sets of mutually unbiased measurements to the quantum-like uncertainty relation.

Lemma 5.11. *If there is a set of mutually unbiased measurements, each of which fully conditionally restricts the state-space; this set of measurements will obey a quantum-like uncertainty relation*

Proof. We prove this by generalising the argument of lemma 5.10. Let $\{X_1, X_2, \dots, X_N\}$ be a set of measurements each of which fully conditionally restricts the state-space. Whenever one of these measurements X_i has an outcome x_j that occurs with certainty, this uniquely specifies the statistics of the other measurements in the set.

This state is unique, but recall that the definition of mutually unbiased measurements allows for the outcomes of any measurement in the set to be permuted to form another valid state. In order for the state to be unique, and for the permutation to take valid state to valid state, the statistics of the other measurements must be described by a distribution that is invariant under permutation: the set where every element is the same. Normalisation then ensures that every outcome probability in this set will be $\frac{1}{N_k}$ where N_k is the number of outcomes to measurement X_k : the distribution with maximum uncertainty.

The same argument can be made for every measurement outcome of every measurement in the set, and hence we see that a set of such measurements obeys a quantum-like uncertainty relation. $\square$

5.7 Trade-off between dynamical and state-space freedoms

We now return to the question of branch locality, and pose the results of the preceding sections in an intuitive, geometric way. Under branch locality, there is a trade-off in between freedoms in the state-space and freedoms in choice of dynamics. On one extreme, for fully independent theories such as box-world, when the position outcome is known¹⁰, there is freedom to choose all of the other statistics, but there are no local operations whatsoever. On the other extreme, for fully conditionally-restricted theories, when the position is fixed such that one outcome happens with certainty, the other statistics are totally constrained, and local operations are possible.

We can visualise this situation geometrically by considering the number of dimensions lost by intersecting the state-space of a system with a hyper-plane of states associated with a particular position outcome happening with certainty. For a d -dimensional theory, the dimensionality of the intersection of a hyper-plane with a fully independent state-space has dimension $d - 1$ (the same as the hyperplane itself). Though there may be some restrictions on the allowed states preventing the spanning of every single probability distribution, there will still be $d - 1$ degrees of freedom in picking a point from the intersection. The rules of branch locality then dictate, for any operation performed on another branch (without support in the chosen branch), not only must the shape of the intersection be invariant¹¹, but every single point in this intersection must also be invariant, leading to the large number of restrictions as seen formally in theorem 5.5.

On the other hand, the intersection with a fully conditionally-restricted theory will be a point (dimension 0), reflecting that once position has been fixed, there are no freedoms remaining whatsoever. Here, as there is only one point in the intersection and it must be mapped to itself (by phase considerations), there are no further constraints on the allowed transformations.

With this in mind, it is also possible to envision intermediate cases, in which some but not all statistics are conditionally restricted (i.e. neither fully conditionally-restricted, nor fully independent) when the position outcome is known. Here, rather than dimensions of freedom in state collapsing to a point, it could collapse into some k -dimensional shape where $1 \leq k < d - 1$. In this scenario, only transformations that act on the freedoms not represented in the intersection will be consistent with branch locality. An example of this partial case will be shown in §5.7.4.

¹⁰We shall take ‘known’ to mean fixed such that the statistics predict one outcome to happen with certainty.

¹¹For reversible transformations, this is already ensured by the fact that transformations are automorphisms of the state-space.

This effect is similar by analogy to the case of body attempting to spin on a surface with perfect friction. If there is a non-zero area of contact between the bodies, then no spinning is possible. However, if the area of contact is reduced to a point, then the friction no longer inhibits the dynamics. (See also the example of the ballerina *en pointe* in [101]).

5.7.1 Example: Qubit interferometer

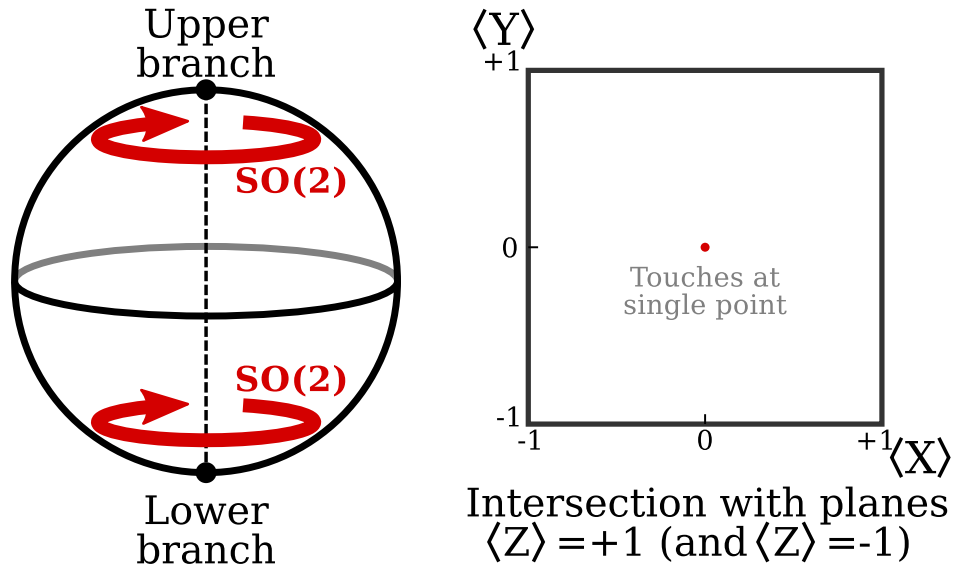


Figure 5.4: Localisable dynamics in a qubit. The full $SO(2)$ phase group can be applied on both the bottom and top branch of a two-level qubit Mach-Zehnder interferometer. LHS illustrates the dynamics as could be implemented through action on the upper or lower branches. RHS illustrates the intersection of the state-space (Bloch sphere) with the planes $\langle Z \rangle = \pm 1$ (by symmetry, they are the same), showing that the sphere touches at just a single point, such that branch locality does not place any further restrictions on the dynamics.

From lemma 5.9, we see that systems described by quantum theory are always fully conditionally-restricted when the position measurement is maximal¹². Thus, from corollary 5.7, we see that the full set of $SO(2)$ dynamics in a two-branch quantum

¹²A subtlety that is important; if there were also an operationally significant “internal” degree of freedom associated with the quantum state, such as spin or polarisation, the position measurement would no longer completely specify the state; and so the phase group of a position measurement over n branches would instead be given by $(\bigoplus^n U(1)) \oplus U(m)$ where m is the dimensional of the internal degree of freedom.

Mach-Zehnder interferometer can be implemented by local operations on either of the branches interchangeably. This is illustrated on the LHS of fig. 5.4.

To visualise the restriction of branch locality on what operations can be applied to the lower branch ($Z = -1$), one must consider the intersection of the state-space with the opposite plane $Z = +1$ that corresponds to the *upper* branch, as it is in states in this plane that have no support in the lower branch. For the quantum Bloch sphere, these two intersections result in the same shape: a single point at $\langle X \rangle = \langle Y \rangle = 0$ (see RHS of fig. 5.4)

For other theories, these intersections could differ such that the dynamics implementable on the top branch will not be the same as those possible on the bottom. This will be seen in some of the subsequent examples.

5.7.2 Example: Stabilizer interferometer

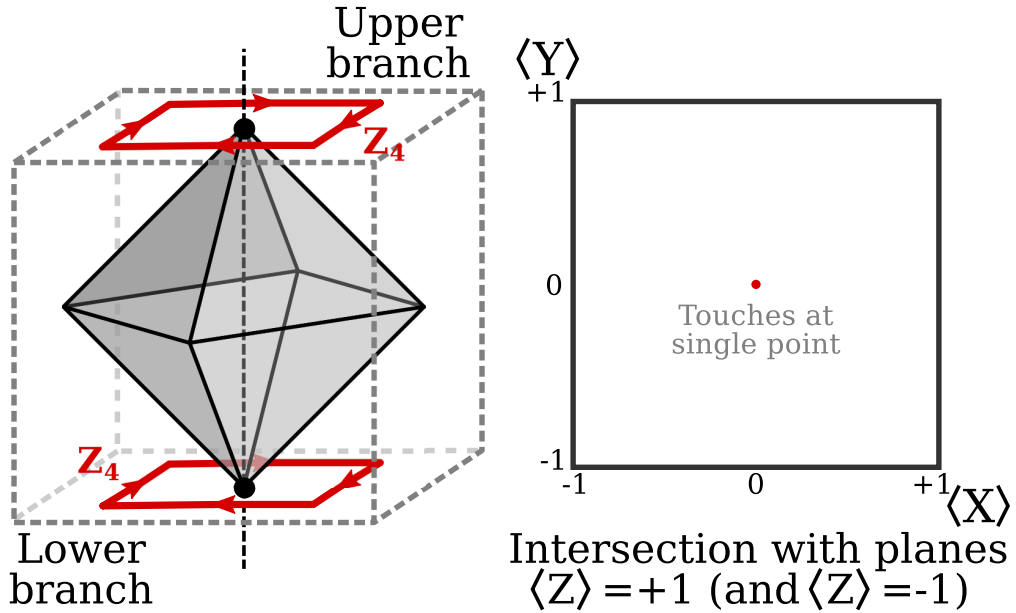


Figure 5.5: Localisable dynamics in a stabilizer-bit. The full phase group of discrete rotations in a square Z_4 can be applied on both the bottom and top branch of a two-level interferometer. LHS illustrates the dynamics as could be implemented through action on the upper or lower branches. RHS illustrates the intersection of the state-space (octahedron) with the planes $\langle Z \rangle = \pm 1$ (by symmetry, they are the same), showing that as with quantum theory, the space touches at just one point, and that branch locality does not further restrict the dynamics.

Consider now the stabiliser subset of quantum mechanics, with probabilistic combinations thereof, as illustrated in fig. 5.5.

The full phase group with respect to the Z measurement is Z_4 , corresponding to rotations in increments of 90° in a square (as stabilizer states are a subset of quantum theory, we also mandate complete positivity on their transformations, ruling out reflections, which would otherwise be automorphisms of the state-space).

This theory satisfies a quantum-like uncertainty relationship, and hence the right-hand side of fig. 5.5 is identical to that of the quantum case (fig. 5.4). By satisfying lemma 5.8 it is fully conditionally restricted, and corollary 5.7 is applicable such that all the phase dynamics can be implemented on either the upper or lower branch. In this way, the theory behaves in exactly the same way as quantum theory in terms of the additional restrictions placed on it by branch locality.

5.7.3 Example: ‘Square-based pyramid bit’ interferometer

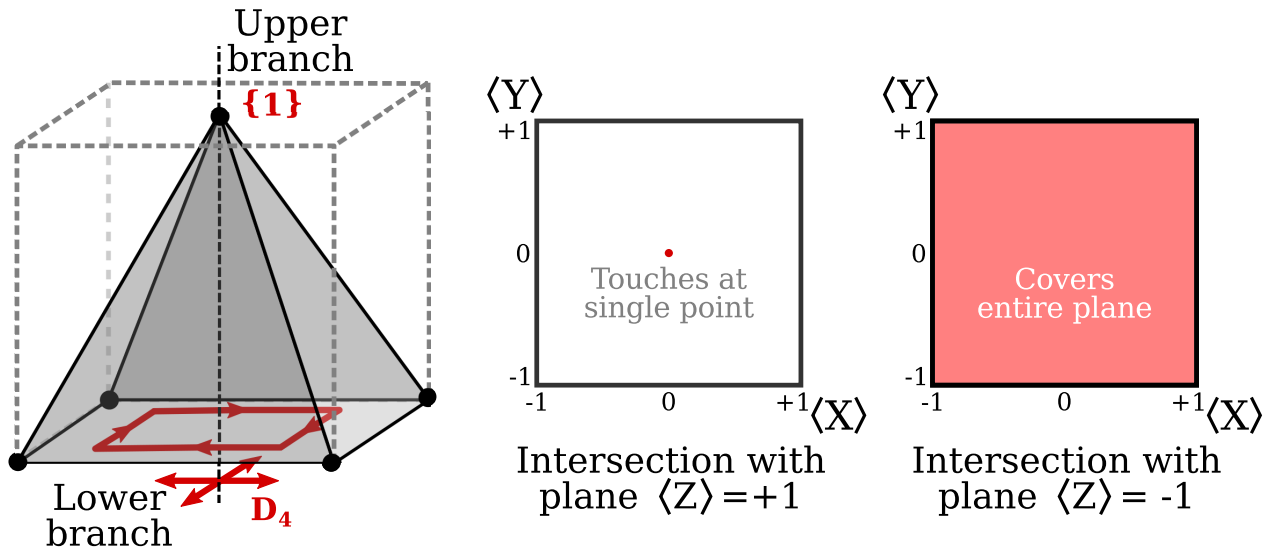


Figure 5.6: Localisable dynamics for a square-based pyramid bit. LHS illustrates the dynamics as could be implemented through action on the upper or lower branches. On the upper branch, no dynamics are possible (as there are states at $Z = -1$ with no support on this upper branch that would be changed). On the lower branch, however, branch locality places no restrictions on the phase operations allowed. RHS illustrates the intersection of the state-space (octahedron) with the planes $\langle Z \rangle = +1$ and $\langle Z \rangle = -1$. For $\langle Z \rangle = +1$, the space touches at just one point, such that lower-branch dynamics are not inhibited. However at $\langle Z \rangle = -1$, the entire surface of states are permissible, totally ruling out the ability to localise dynamics to the upper branch.

Now let us consider a very asymmetric example. Take a theory with three binary measurements (one of which represents position) where when the position measurement suggests the particle to be in the upper branch with certainty ($Z = +1$), the

other two measurements must be maximally uncertain (giving either $+1$ or -1 with equal probability); but when the position measurement suggests the particle is in the lower branch with certainty ($Z = -1$), the other two measurements are free to take any statistics.

The convex hull of these two cases results in a state-space given by a square-based pyramid, as illustrated in fig. 5.6. This theory is clearly neither fully independent nor fully-conditionally restricted (but rather, a mixture of both depending on the Z statistics), but its behaviour under branch locality can be well understood by the geometric arguments made in this section.

In absence of any other constraints, the full phase group of this state-space will be the automorphisms of the square (drawing from the symmetries present in any plane of fixed Z). For actions taken on the top branch of the interferometer, to probe the constraints of branch locality we must consider the states which have no support on the top branch: the states with support only on the lower branch. There are two dimensions of freedom in selecting a state on this branch, and from this we can create a sufficient number of independent constraints from branch locality to totally restrict the set of locally-applicable dynamics to the trivial group $\{\mathbb{1}\}$ ¹³. On the other hand, for operations taken on the lower branch, there is but one unique state without support (the tip of the pyramid), and under branch locality considerations, this places no additional constraints than were already present in the phase group (by similar arguments made in theorem 5.6). Thus any transformation in the phase group can be implemented by action taken on this branch.

This state-space therefore makes for a very odd interferometer, where the interference patterns (§ 4.2) can only be induced by actions on one preferred branch. This asymmetric behaviour should come as no surprise, as it follows from the asymmetry of the state-space. Moreover, the shape of this state-space makes it somewhat hard to describe a coherent beam-splitter, and so this particular example likely only has use as a toy for considering the extremes of behaviours possible under branch locality!

5.7.4 Example: Spekkens' toy model interferometer

Consider now an interferometer where the position measurement is encoded by the Z measurement ( vs. ) in Spekkens' toy model [28, 29] (see § 2.7). Without the epistemic restriction, the state-space corresponds to the tetrahedron drawn in fig. 5.7,

¹³ As the total phase group is finite, we can prove this by exhaustion. For every element in the phase group except for $\mathbb{1}$, one can pick a point on the lower surface which according to branch locality the statistics of which should not change under the application of an operation on the upper branch, but in fact is changed. Thus every operation but $\mathbb{1}$ can be ruled out.

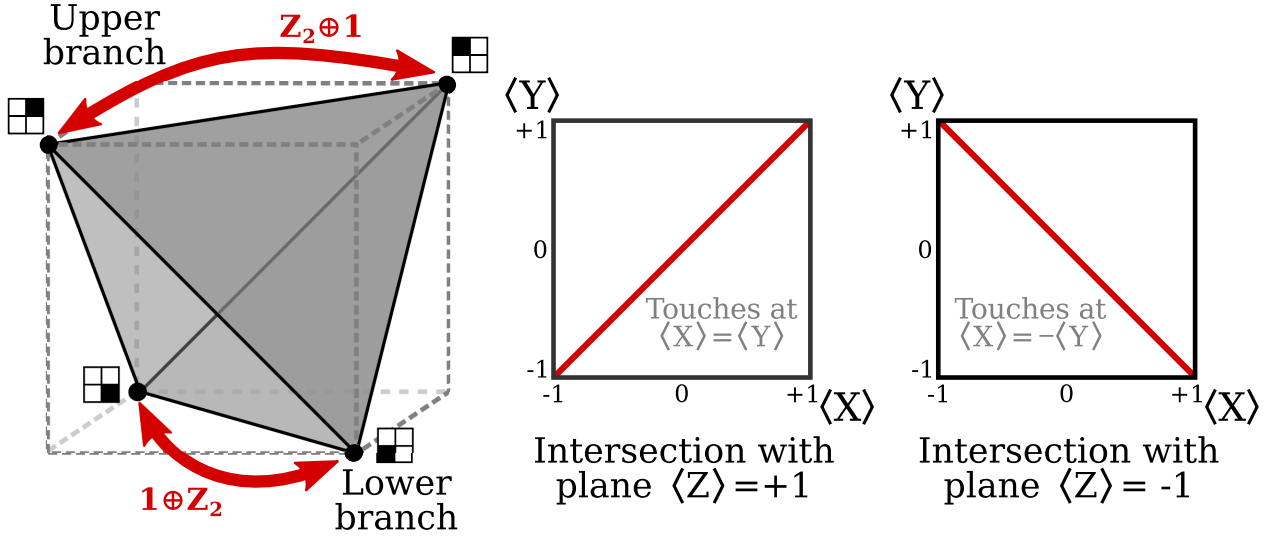


Figure 5.7: Localisable dynamics for ontic Spekkens'-bit. LHS illustrates the dynamics as could be implemented through action on the upper or lower branches. On the upper branch, the permutation $(12)34$ is possible, corresponding to a flip in one direction. On the lower branch, the permutation $12(34)$ is possible, corresponding to a flip in the other direction. RHS illustrates the intersection of the state-space (octahedron) with the planes $\langle Z \rangle = +1$ and $\langle Z \rangle = -1$. In both cases, the freedoms are reduced to a line, such that some operations are ruled out by branch locality, and some remain possible. However, the line is in orthogonal directions for the two planes, such that even though not all the operations possible from either branch, between the two branches, the entire set of transformations is possible.

and the allowed reversible phase transformations correspond to the automorphisms of this state-space, the permutations $\{1234, 2134, 1243, 2143\}$ (isomorphic to group $Z_2 \oplus Z_2$).

In this case, the state-space is neither fully independent nor fully conditionally restricted, and this results in an interesting division of the dynamics. First, consider the dynamics that may be localised to the top branch. The states without support on this branch, are in the plane $Z = -1$, which admits every state where X and Y are perfectly anti-correlated. The permutation 2134 which exchanges X with $-Y$ (and also $-X$ with Y) has no effect on any point on this line. (Another way to visualise this invariance is to note that the transformation only swaps possibilities 1 and 2 and a state on this surface only has support on 3 and 4 which are not affected by this). On the other hand, the permutation 1243 (exchanging X with Y) will map states on this line to states with measurably different statistics, when branch locality would otherwise forbid it. Thus, only the subgroup of permutations $\{1234, 2134\}$ (alternatively written $Z_2 \oplus \mathbb{1}$) can be localised to the top branch.

Similar arguments can be made for operations localisable to the lower branch. However, here when one considers the intersection with $Z = +1$, one is restricted to

the line of states where X and Y are perfectly correlated. The outcome is that only permutations in the subgroup $\{1234, 1243\}$ (written $\mathbb{1} \oplus Z_2$) may be implemented locally on the lower branch.

Thus, we see for the tetrahedral ontic state-space, although the full group of phase transformations may be implemented over the whole system, each branch has access to a different subset of transformations.

Suppose now we enforce the epistemic restriction on the state-space. This restricted state-space is now the octahedron embedded within the tetrahedron (by taking the convex closure of the six midway points of each edge). Crucially now, the theory obeys a quantum-like uncertainty relation, and is fully conditionally-restricted. The full set of phase dynamics $Z_2 \oplus Z_2$ may now be implemented from either the top or the bottom branch of the interferometer. The ‘censorship’ of the information about the ontic state, functioning like an uncertainty relationship, has enabled more dynamics on each branch. However, the overall set of dynamics possible on either branch remains the same.

This curiosity, in which an ‘ontic’ state seemingly violates locality considerations but is ‘saved’ by the inability to ever measure this quantity to full precision appears in another famous hidden-variable theory: the de Broglie-Bohm pilot-wave theory of quantum mechanics (see e.g. [102–106]). In such a theory, it is possible [106] to construct a set-up involving an entangled pair of particles entering Stern–Gerlach set-ups at two different positions in space, such that if the exact initial position of the one of the particles was known, at one detector it would be possible to determine from the final trajectory taken by the particle whether or not a measurement had made by the remote detector. This would thus violate non-signalling (i.e. one of the strongest ‘locality’ conditions). However, in the theory of de Broglie-Bohm mechanics, knowing this initial position is forbidden, and the violation is avoided by epistemic restriction.

5.8 Deutsch’s algorithm beyond quantum theory

Let us now return to Deutsch’s algorithm [21, 22], previously mentioned in § 5.1. In order to implement this in an interferometry set-up (as illustrated in fig. 5.1), we require there to be a choice of at least two locally applicable operations (i.e. $\{\mathbb{1}, Z\}$) at the upper branch, as well as two operations locally applicable at the lower branch, such that either operation in isolation has the same measurable final effect, and the application of both has no effect at all. This required behaviour ($0 + 0 = 1 + 1 = 0$; $0 + 1 = 1 + 0 = 1$) means that we must be able to implement elements from Z_2 as transformations localised to each of the branches in order to perform the algorithm.

We can now see directly which theories will allow us to do this, and which will not.

In quantum theory, it is possible because $Z_2 \subset \text{SO}(2)$ (choosing elements $\{0, \pi\}$), and $\text{SO}(2)$ can be applied from either of the branches.

Likewise, we can implement Deutsch's single-bit algorithm using the stabilizer subset of quantum theory, as $Z_2 \subset Z_4$ (which is itself a subset of $\text{SO}(2)$) and Z_4 can be applied from either branch. It is not a problem that there are operations in Z_4 that we would not allow for the purpose of this algorithm, or that there are multiple ways of selecting transformations that recover the desired multiplication table (e.g. taking either $\{+0^\circ, +180^\circ\}$ or $\{+90^\circ, +270^\circ\}$), as this was also the case in quantum theory, where any two phase parameters that differ by $(2n + 1)\pi$ for integer n would have achieved the same output statistics.

On the other hand, for a 3-in 2-out gbit interferometer where the position measurement does not fix the other statistics, the only locally implementable operation is the identity. Thus, it is not possible to perform Deutsch's algorithm with such a set up. Likewise, the pyramid toy model in § 5.7.3 (fig. 5.6) would not work, as it is only possible to implement Z_2 on one of the two branches, so it is impossible for the output of the device to be a comparative function of some unknown property in each branch.

More subtle is the example of Spekkens' toy model. For the ontic tetrahedral state-space, the division of dynamics into $Z_2 \oplus \mathbb{1}$ for one branch and $\mathbb{1} \oplus Z_2$ for the other means that the required structure cannot be satisfied. Even though dynamics can be implemented on both branches, they do not interact with each other (behaving instead exactly like actions taken on two separate classical bits). Indeed, to reliably measure one outcome for $\mathbb{1} \oplus Z$ and $Z \oplus \mathbb{1}$ and another for $\mathbb{1} \oplus \mathbb{1}$ and $Z \oplus Z$ would require a device that could tell the difference between eigenstates of the X measurement $\left\{ \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}, \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \right\}$ and eigenstates of the Y measurement $\left\{ \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}, \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \right\}$. This may be seen by taking the state after the first beam-splitter to be any allowed state except for $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ or $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ (if it were one of these states, then one of the branches could have no effect on the state regardless of the choice of operation), and considering what the state is following the application of an element of $Z_2 \oplus Z_2$. Such a device cannot be constructed for a single bit in Spekkens' toy model (this can be demonstrated by exhaustively checking every possible permutation and seeing that, as with linear quantum mechanics, there is no operation that can map two non-orthogonal states into orthogonal states). Thus, the tetrahedral state-space is not suitable for implementing Deutsch's algorithm.

However, when we take the full toy model, including the epistemic restriction, nothing rules out both branches choosing to implement an element of $Z_2 \oplus \mathbb{1}$ (or equivalently, both from $\mathbb{1} \oplus Z_2$), and once again the algorithm can be implemented with similar statistics as in the quantum case. This example illuminates the explicit importance of the uncertainty principle in enabling interesting dynamics in an interferometer in Spekkens' toy model. With the uncertainty constraint on the state-space arising from the epistemic restriction, Deutsch's algorithm is possible; without it, it is not.

5.9 Extensions to branch locality

Finally, we note that whilst the postulate of branch locality combined with the requirement for non-trivial dynamics motivates the need for an uncertainty restriction, it does not recover the entire quantum state-space. The restriction of branch locality, that a transformation applied to a branch should have no effect on states with *no* probability of support on that branch, can be seen to be slightly too weak from a couple of counter examples.

First, consider the state-space illustrated in fig. 5.8: a ‘trimmed sphere’ such that when $\langle Z \rangle = \pm 1$, then $\langle X \rangle^2 + \langle Y \rangle^2 \leq \epsilon$ for some small $\epsilon > 0$. Suppose no such trimming occurs for the extremities of $\langle X \rangle$ and $\langle Y \rangle$, such that the state-space retains circular symmetry about the Z axis, and the full phase group of the system is given by $SO(2)$.

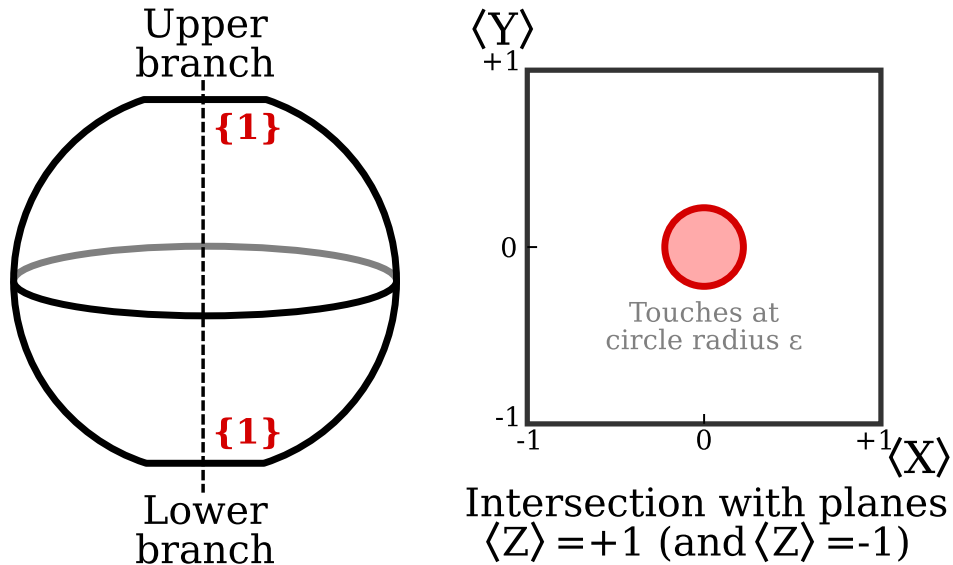


Figure 5.8: Counter-example state-space with no localisable dynamics. *LHS* shows the ‘trimmed sphere’ state-space. *RHS* illustrates the intersection of the state-space with the planes $\langle Z \rangle = +1$ and $\langle Z \rangle = -1$ forming a small circle of allowed states. The freedom in choice of state on these planes, for any positive ϵ is sufficient to rule out the localisation of any phase dynamics, despite the fact the state-space is almost that of a qubit.

For any $\epsilon > 0$, this theory is technically fully independent: two degrees of freedom remain in choosing the state even after fixing $\langle Z \rangle = \pm 1$. For this reason by theorem 5.5 (or by the geometric argument in § 5.7), we see that *no* dynamics can be localised to either the upper or lower branch of an interferometer when the particle is described by such a state-space. Moreover, we can make ϵ an arbitrarily small positive number, and this will still be the case. It is only at the limit of $\epsilon = 0$, where the state-space becomes exactly that of the qubit that dynamics are possible without violation of branch locality.

That two state-spaces that can be made arbitrarily close to each other should have such widely different dynamics illustrates the shortcoming of the definition of branch locality.

The second example is to consider theories with a state-space subject to the constraint

$$|\langle X \rangle|^n + |\langle Y \rangle|^n + |\langle Z \rangle|^n \leq 1 \quad n \geq 1, \quad (5.53)$$

and note that for large (but non-infinite) n , the state-space converges on that of a cubic gbit, as illustrated in fig. 5.9. Except when $n = 2$ (where the example becomes exactly the qubit), this state-space will have square symmetry about the Z axis, and so in the absence of other constraints has the same group of reversible phase transformations as a cubic gbit, D_4 . However, despite the fact this state-space can be made arbitrarily close to that of a gbit, is in fact fully conditionally restricted. This is because for any finite n , when $\langle Z \rangle = \pm 1$, the only allowed statistics for the remaining measurements are $\langle X \rangle = \langle Y \rangle = 0$. Hence, it is subject to theorem 5.6, and corollary 5.7: all of the dynamics can be implemented locally on either of the two branches.

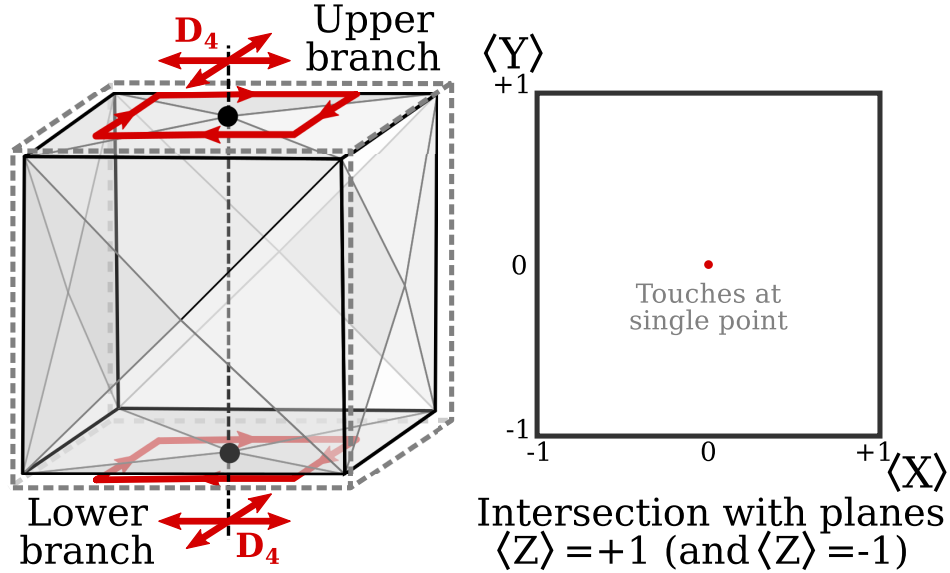


Figure 5.9: Counter-example state-space with localisable dynamics. *LHS shows the ‘almost cubic’ state-space, where the surface has been pinched such that it almost fills the cube, but still only has a single state associated with each certain measurement outcome. RHS illustrates the intersection of the state-space with the planes $\langle Z \rangle = +1$ and $\langle Z \rangle = -1$, which is reduced to a point. Thus, the theory is still fully conditionally-restricted and every transformation can be localised to either branch, despite the fact that the state-space is almost that of a gbit.*

Thus, we see that this state-space and that of the cubic gbit can be made arbitrarily close to each other, but the branch locality restriction still affects them differently.

These problematic examples arise because definition 5.1 of branch locality imposes restrictions on a limiting case (total lack of support), without postulating any behaviour for the system as it converges upon this limit. This allows us to construct the above counter-intuitive examples where the behaviour changes from fully independent to fully conditionally restricted (or vice versa) only at the asymptote.

Thus, it may be desirable to patch the definition of branch locality to also impose restrictions on the behaviour as it tends towards states without support:

Definition 5.6: Strong branch locality

As the probability of a state being on a branch is reduced towards zero, the measurable effect of applying an operation locally to that branch must also diminish. If there is no support, the operation should have no effect.

The strong definition of branch locality could censor the dynamics in the almost-cubic state-space, as in that example there are states with statistics close to $\langle Z \rangle = 1$ which when acted on by an operation such as a flip applied at $Z = -1$, changes the X statistics from almost 1 to almost -1 . This would be a very noticeable change for states with very little support on the branch in which the operation was applied.

However, this definition is much harder to quantify than the simple version of branch locality. In definition 5.1, both the concept of support and of ‘measurable effect’ were binary conditions: either the state has some support, or it has no support; and either there are some statistics that change, or none of them do. Whilst level of support can naturally be quantified by the probability of measuring it in a position, it is less straightforward how to quantify the ‘noticeability’ of a transformation. Moreover, it is not obvious how best to motivate the exact trade-off between support and ability to change the state in a way that would both be consistent with quantum theory, but not a priori lead to it.

One plausible way to quantify the noticeability of a transformation is to consider the visibility of interference fringes. Take the example in quantum theory of a two level Mach-Zehnder interferometer. Let the state after the initial beam-splitter be given by $|\psi\rangle = \sqrt{1-p}|0\rangle + \sqrt{p}|1\rangle$ (for non-negative real p) and the phase operation $U = \text{diag}(1, e^{i\phi})$ is applied, one ends up with the state $|\psi'\rangle = \sqrt{1-p}|0\rangle + e^{i\phi}\sqrt{p}|1\rangle$. Calculating the statistics of this state, one sees that the influence of ϕ can be measured by making an X measurement, such that $\langle X \rangle = 2\sqrt{p(1-p)}\cos\phi$. As $p \rightarrow 0$ (reducing support on the branch in which the transformation is applied), the amplitude of the influence of ϕ is reduced, and at $p = 0$, $\langle X \rangle = 0$ and the choice of ϕ is totally undetectable in the statistics of the state. This behaviour is qualitatively consistent with definition 5.6.

We cannot require a monotonic relationship between support and ‘observability of

change' over the entire range of $p \in [0, 1]$, if we want a definition of branch locality that is consistent with quantum mechanics. In this example, when $p > \frac{1}{2}$, increasing p also reduces the visibility of the fringes (fringes are maximally visible for $p = \frac{1}{2}$).

Thus, we leave as an open question what, mathematically, the more precise physically-motivated definition of strong branch locality should be.

Finally, we note that the almost-qubit example still falls afoul of the updated definition of branch locality. Its behaviour as it approaches the extremes from a distance satisfies the definition: when there is little to no support on the branch, the circle of possible states is small, and so the statistics do not vary by very much. However, at and very near the limit of no support, permitting dynamics to be localised yields incorrect behaviour because minute detectable changes can be affected on branches with no support whatsoever.

This is not as much of a problem as it might seem: the existence of such a state-space would require a universe that permitted some degree of violation of branch locality. This makes for a sufficiently drastic difference between the counter-example, and the true Bloch sphere, that it is reasonable to expect different behaviours, despite the state-spaces being otherwise similar. Suppose we attempted by experiment to tomographically map all allowed states in the Bloch sphere. Due to the finite precision intrinsic to any experiment, we could not be sure if we had the perfect Bloch sphere, or the clipped sphere as in fig. 5.8. If we believe branch locality should hold, we would then be motivated to opt for the truly spherical state-space over the nearly-spherical alternative.

An alternative may be to impose some fundamental limit on the precision of a single measurement, such that there is ultimately some ϵ from which it will never be possible to distinguish $\langle X \rangle = \epsilon$ and $\langle X \rangle = -\epsilon$. In this theory, the small ϵ clipped-sphere, and the true Bloch sphere would be indistinguishable. However, such a theory would be intrinsically non-convex, and hence beyond the scope of description by the GPT framework.

5.10 Conclusion

In this chapter, we introduced an operational definition for locality in the context of branching interferometers: if a system has no probability of being in a particular branch, then actions applied on that branch should not affect any of the statistics of the system. This definition proved to place extreme restrictions on which transformations can be localised to any subset of the interferometer, when there are not enough restrictions on the state-space. However, we found that by imposing constraints such as the uncertainty principle on the state-space, interesting dynamics are

once again possible. This highlighted the important enabling role that uncertainty plays in an interferometer. By applying this in a more computational context, we determined which theories could implement the Deutsch-Jozsa algorithm [21, 22, 26] in an branching interferometer: quantum theory and stabilizers bits can, whereas gbits can not. Spekkens' toy model proved to be an interesting case, in that it can only perform Deutsch's algorithm when the epistemic principle (itself a type of uncertainty condition) is in place.

On the other hand, we saw that the definition for branch locality is not quite strong enough to fully recover the state-space of the quantum bit. We suggested that a good definition must also account for the behaviour of a system as it approaches the limit of no probability on a branch, but it remains an open question exactly how to phrase this in the most mathematically useful way.

In the next chapter, we will consider the restriction placed by a different kind of locality on the behaviour of generalised interferometers, which arises by considering *relativity of simultaneity* [30].

Chapter 6

Special relativity and interferometers

Relativity is the physical concept that all observers should witness the same laws of physics, even when located at possibly different positions, facing in different directions, or moving at different speeds. In classical Galilean relativity (see e.g. [107]), two observers observing a system would both see that system evolve according to the laws of Newtonian physics, even if they were moving at different constant speeds. Moreover, they would not be able to tell if they themselves were moving or standing still just by watching that system.

In his seminal 1905 paper, Einstein reconciled the concept of relativity with classical electromagnetism by suggesting that the speed of light also has a constant value, regardless of the speed of the observer who measures it [30, 108]. This concept, known as *special relativity*, which concerns itself with the observations made by observers in *inertial reference frames* (non-accelerating, non-rotating points of view) has wide-reaching consequences. Over and above Galilean relativity, in which observers may disagree on the position or speed of a witnessed object, in special relativity observers may also disagree on the length of an object, and the amount of time between two events happening to that object. When two events happen in different positions in space (i.e. are separated by a *space-like interval*), observers moving at different speeds may even disagree on the order in which these events happen, in a phenomenon known as the *relativity of simultaneity*.

In this chapter¹, we shall first consider which parts of the GPT framework should not change when viewed by different observers (§ 6.1). We will then consider the consequences that arise in the context of a Mach–Zehnder interferometer, when we require it to be consistent with relativity of simultaneity (§ 6.2). Next, we shall discuss the possibility of simulating non-quantum interferometers that are compatible with special relativity (§ 6.3). Finally, we will consider a family of theories in the GPT framework that describe two-level systems (n -balls), and see how the physically motivated results in this chapter lead towards quantum theory as a reasonable description of reality (§ 6.4).

¹Some results of this chapter were presented in [3].

6.1 Relativity and the GPT framework

As an operational framework, generalised probabilistic theories are primarily concerned with measurement statistics, which might correspond to clicks in a detector. Consider an experiment that is run just once, consisting of a detector that has the possibility to click. If there is a click, the device is configured such that it triggers sequence of events, that could be arbitrarily drastic (e.g. triggering a cascade in a detector, flashing a light, or setting off an explosion). Two observers, differing only in their speed or relative location, should not disagree about whether the device has triggered or not: the existence of the event corresponds to an objective reality, even if they observers otherwise disagree on where and when it happened.

The probability of a detector clicking, either in the interpretation corresponding to relative frequencies of outcomes, or as weightings for a rational betting strategy, should also not change between reference frames. In the GPT framework, probabilities arise by taking the inner product of a state s with an effect e , such that the probability of measurement outcome associated with e is given by $P(e | s) = e \cdot s$. Thus, these inner products $e \cdot s$ should not change between reference frames, and hence are Lorentz invariant². The statistics given by $e \cdot s$ can be changed by acting on the state with a transformation T , such that $P(e | Ts) = e \cdot Ts$, and this quantity should also be Lorentz invariant.

Let us now consider explicitly how T might change under a change of reference frame. Suppose there is a reference frame change $\alpha : R \rightarrow R'$, and another $\beta : R' \rightarrow R''$. Let T, T' and T'' correspond to the same physical transformation observed in R, R' and R'' respectively. Let the map Ξ_α take T into $T' = \Xi_\alpha(T)$, and likewise Ξ_β take T' into $T'' = \Xi_\beta(T')$. One can form an equivalent map that takes R straight to R'' as $\beta\alpha : R \rightarrow R''$, and one should also be able to map T to T'' with a compound operation $\Xi = \Xi_\beta\Xi_\alpha$. For any R and R'' , there is an infinite choice of α, β and R' such that $R \xrightarrow{\alpha} R' \xrightarrow{\beta} R''$. In all cases, the numerous possible Ξ_α and Ξ_β must always combine to form exactly the same Ξ that takes T to T'' . For this reason, the maps $\{\Xi\}$ on the transformations must be a representation of the Lorentz group: they combine in such a way to give exactly the same multiplication tables.

We will go one step further, and assume specifically that *transformations* on state-spaces encoding the “which branch?” freedom in an interferometer do not change between reference frames. The motivation for this arises from the the type of freedom

² The mathematical group representing the change in observable properties following the rotation or boost (change in velocity) of an observer is known as the Lorentz group. If translations in space and time are also included, one has the Poincaré group. Quantities that remain constant when viewed in different reference frames are called *Lorentz invariant*, as acting on them with a member of the Lorentz group effects no change.

described by the state-space of the interferometer. The “which branch?” measurement discriminates between unambiguously labelled alternatives (i.e. “in branch A ” or “in branch B ”), as opposed to some *geometric* freedom that could change under rotation or boost (e.g. the direction which the spin of a particle points in). Even if two observers disagree about which of the branches is on top of the other (such as would be the case if one observer is orientated at 180° to the other), they can identify which branch is which by some other labelling. As such, if one constructs the state of the system explicitly in the manner of eq. (2.11) by stacking probabilities on top of each other, then the probabilities P (in branch A) and P (in branch B) will be the same in all reference frames. It then follows that if some physical action T is performed on the system and possibly alters the probabilities of being in one branch or another, then all observers must agree on how T changes the statistics associated with the “which branch?” measurement, otherwise it would lead to disagreement about the final statistics.

This reasoning can be extended to the other measurements in the state-space, by also considering the choice of measurement to be something that all observers agree on. Suppose one had, for example, one of the three devices illustrated in fig. 2.3 which measure X , Y or Z , and these three devices could all be unambiguously distinguished. It could be taken that which of these devices is attached to the experiment (and which detector in the device clicks) is agreed upon by all observers, regardless of their reference frame (in contrast to a Stern-Gerlach device whose orientation in space is an intrinsically geometric property, and could be seen differently by observers in different reference frames). In this set-up, not only would the “which branch?” statistic be agreed upon, but every statistic would be the same between reference frames. It then follows that T must act on these statistics in the same way in every frame.

Thus, in this context, the action of the Lorentz group on T can be taken to be trivial. We have argued for this in a specific interferometric setting, with unambiguous measurements and outcomes. However, this regime laid out here is not the most general one; it would be particularly ill-suited to describe the location or momentum of a particle in free-space.

6.2 Relativistic behaviour of interference experiments

Consider a spatially disjoint interferometer (such as the Mach-Zehnder interferometer) with two branches. Let there be independently acting agents (Alice and Bob) on each branch, who at the extreme point of each branch can choose which action (if any at all) they wish to take by inserting a choice of phase plates into the device. The action of the agents, as discussed in the preceding chapters, constitutes a transformation in the *phase group* of the position measurement. As illustrated in fig. 6.1, we will show

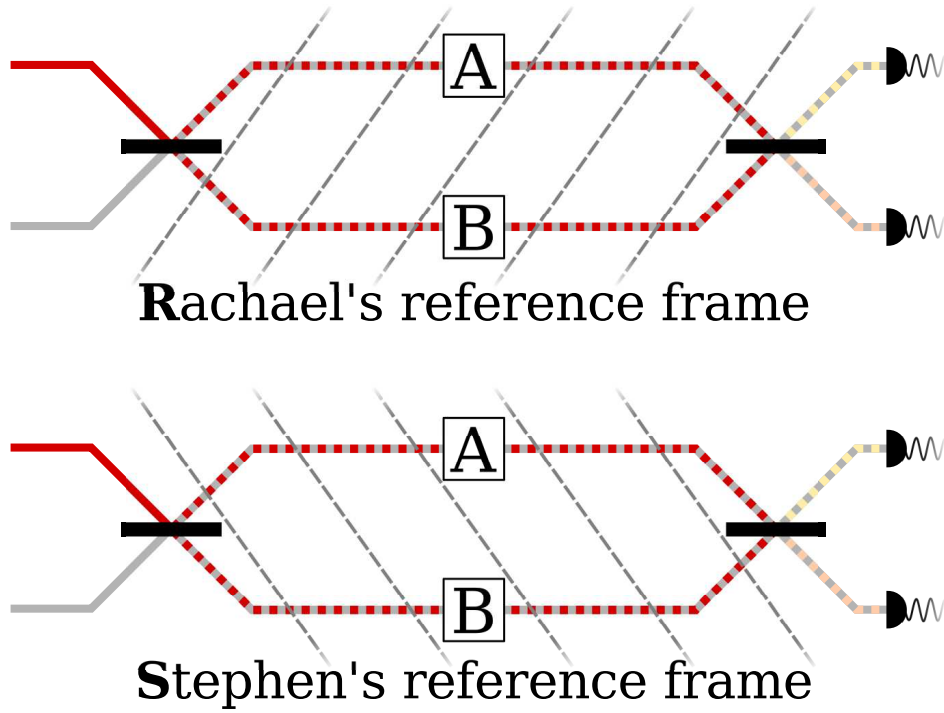


Figure 6.1: A Mach-Zehnder interferometer viewed from two reference frames. The dashed diagonal grey lines correspond to events judged as simultaneous by an observer. Rachael (top) hence observes particles arriving at phase plate A before phase plate B. On the other hand Stephen (bottom) observes particles arriving at phase plate B before phase plate A. If the phase operations associated with A and B do not commute, the two observers might disagree about the final statistics (i.e. interference pattern) output from the interferometer.

that relativistic considerations place a further restriction on the phase dynamics³ that Alice and Bob can apply.

Lemma 6.1. *Phase transformations applied to spatially disjoint branches of an interferometer must commute, if they are to be consistent with relativity of simultaneity.*

Proof. Let the state after the first beam-splitter be given by s , the action taken by Alice be T_A , and the action by Bob be T_B . If there is some choice of T_A and T_B such that $[T_A, T_B] \equiv T_A T_B - T_B T_A \neq 0$, then the order in which the operations are applied will be significant as in general $T_A T_B s \neq T_B T_A s$, and there will hence be some measurable difference in the output statistics.

Suppose the order in which the operations are applied is given by when the particle (should it be in one branch) first encounters the phase plate. If Alice's plate is encoun-

³Throughout this chapter, we shall use 'phase dynamics' to implicitly refer to phase dynamics relative to the position measurement Z .

tered before Bob's, then the output statistics will be $T_B T_A$ s, and if Bob's is encountered before Alice, $T_A T_B$ s. However, as illustrated in fig. 6.1, these phase plates are applied at spatially-disjoint locations, and so it is possible to choose two observers ('Rachael' and 'Stephen') who disagree on the order in which the phase plates are encountered.

In Rachael's frame, the A happens before B and the operation is given by $T_B T_A$. On the other hand, in Stephen's frame B happens before A and the operation is given by $T_A T_B$. Rachael's frame and Stephen's frame differ by a Lorentz transformation, and we have assumed that transformation on an interferometer's state-space should be Lorentz invariant, including the compound operation formed by applying T_B after T_A . Thus

$$T_A T_B = T_B T_A. \quad (6.1)$$

If the phase operations T_A and T_B do not commute, such that $T_A T_B \neq T_B T_A$, then there will be a contradiction. $\square$

If one were disinclined to make the assumption (discussed in § 6.1) that transformations on the state-spaces of interferometers do not change between reference frames, then the conclusion of this lemma, $T_A T_B = T_B T_A$, could in itself be taken as a reasonable assumption instead. One way to motivate this would be to consider how actions taken in two different locations (A and B) in a bipartite system are treated in a non-signalling theory. An action T_A taken on A should not immediately affect any statistics of the part of the system in B . Then by considering what might happen to various product states, one concludes that the action should be written $T_A \otimes \mathbb{1}$. Likewise, for actions taken on B , one would expect the effect on the entire system to be $\mathbb{1} \otimes T_B$. These two operations naturally commute (conversely, one could start with commutativity between actions applied to subsystems and deduce no-signalling, as per assumption 4 and corollary 1 in [18]). This behaviour can then be extended into the assumption that actions on different branches within a single system behave in the same way⁴, and must also commute.

This element-wise restriction will naturally limit the set of allowed transformations under certain conditions. We shall make the following definition for conciseness:

Definition 6.1: Symmetrical interferometer

A (two-level) interferometer is **symmetrical** if the sets of phase transformations that can be applied on either branch, $\{T_A\}$ and $\{T_B\}$, are identical.

It is important to stress this definition requires the sets be identical rather than *isomorphic*. For example, an interferometer with the phase groups $\{T_A\} = \mathbb{1}_3 \oplus \text{SO}(3)$

⁴The link between bipartite systems, and two-branched systems is particular obvious in standard quantum theory where one can convert between first and second quantized descriptions of an interferometer using $|\text{in } A\rangle \leftrightarrow |1\rangle_A \otimes |0\rangle_B$ and $|\text{in } B\rangle \leftrightarrow |0\rangle_A \otimes |1\rangle_B$.

and $\{T_B\} = \text{SO}(3) \oplus \mathbb{1}_3$, would not be considered symmetrical even though there is an obvious isomorphism between T_A and T_B .

Corollary 6.2. *In a symmetrical interferometer, the locally applicable phase dynamics must be abelian in order to be consistent with relativity of simultaneity.*

Proof. If $\{T_A\} = \{T_B\} = \{T\}$, for every a, b in $\{T\}$, following from lemma 6.1 $[a, b] = 0$ and hence $\{T\}$ is abelian. $\square$

Corollary 6.3. *If a two-branch interferometer is described by a theory that is fully conditionally-restricted with respect to the position measurement, and the only restriction on the transformations come from branch locality, then the total phase dynamics must be abelian.*

Proof. Corollary 5.7 tells us that in a fully conditionally-restricted theory without other restrictions, every element of the phase dynamics can be applied locally on either branch and thus the interferometer is symmetrical. From corollary 6.2, the associated phase group must be abelian. $\square$

It so transpires that there is only a limited number of distinct groups that can both correspond to transformations within the GPT framework and that are also abelian. As the group of reversible transformations in the GPT framework correspond to the automorphism group of a convex set (or some subset of these automorphisms), these groups will necessary be either finite or compact. The well-established mathematical properties of these types of groups allow us to make a rather restrictive statement on what phase dynamics are compatible with relativity.

Theorem 6.4. *The phase group $\mathcal{G}_\Phi$ of reversible dynamics in a symmetrical interferometer, if it is to be consistent with relativity of simultaneity, is necessarily isomorphic to a group $\mathcal{G}_{\text{ab}}$ where $\mathcal{G}_{\text{ab}} \subseteq \bigoplus^n \text{SO}(2)$.*

Proof. Corollary 6.3 tells us that symmetrical interferometers must have abelian phase dynamics if they are to be compatible with relativity of simultaneity.

First, we address finite groups. The fundamental theory of finite abelian groups (see e.g. §11.12 in [109]) tells us that that every finite abelian group can be expressed as some direct sum of cyclic groups (i.e. $Z_{n_1} \oplus Z_{n_2} \oplus \cdots \oplus Z_{n_m}$). It can be further seen that each $Z_n \subset \text{SO}(2)$ by discretising the allowed rotation angles in $\text{SO}(2)$. (Visually, this is like drawing a regular polygon with n edges inside a circle). Thus, finite abelian groups are all a subset of $\bigoplus \text{SO}(2)$. The number of $\text{SO}(2)$ addends required to form the superset may in general be less than the number of cyclic addends, as if n and m are co-prime, then $Z_n \oplus Z_m \cong Z_{nm}$, and hence $Z_n \oplus Z_m \subset \text{SO}(2)$.

A similar concept known as *maximal tori* holds for compact abelian Lie groups. Theorem 5.2 in [110] tells us that every compact abelian Lie group is isomorphic to

$$\mathcal{G}_{\text{ab}} = \bigoplus^m \text{SO}(2) \oplus F, \quad (6.2)$$

where F is a finite abelian group (and could be $\mathbb{1}$ if $\mathcal{G}_{\text{ab}}$ is connected⁵). As these finite abelian groups have $\bigoplus^n \text{SO}(2)$ as a superset, we thus see that all abelian dynamics can be expressed

$$\mathcal{G}_{\text{ab}} \subseteq G_{\text{torus}} = \bigoplus^n \text{SO}(2). \quad (6.3)$$

Thus, the phase group $\mathcal{G}_{\Phi}$ must be isomorphic to the group $\mathcal{G}_{\text{ab}} \subseteq \bigoplus^n \text{SO}(2)$. $\square$

6.3 Simulating interferometers

Let us consider the question of *simulating* a two-branch interferometer. The particulars of simulation depend on the experiment being emulated, so in this instance we will consider an experiment where an interferometer is used to probe an unknown sample. As discussed in chapter 4, this sort of experiment involves the identification of an element in a phase group by the observation of output statistics, and so we are specifically interested in simulating the dynamics (rather than attempting to cover the entire state-space of the device). Crucially, the MZI experiment with a sample on just one branch reveals an element of a phase group that was not a priori known to the experimenter.

A simulation of this experiment, using another interferometer (or a set of many interferometers), is thus a game where one player knows what the hidden element of the group is in advance, and communicates this to another player by applying some phase operation on the interferometer they have. (If this is done with exactly the same type and number of interferometers as the set-up being simulated, it would arguably be a run of the experiment itself, rather than a simulation!) Thus, we see that simulation relies on the existence of an epimorphism⁶ from the dynamics of the device being simulated onto the dynamics of the simulating device.

We will show (by construction) that if an interferometer has a phase group compatible with relativity of simultaneity, it is possible to build a simulation of this experiment

⁵Connectedness will be discussed further in §6.4.1.

⁶This does not need to be strictly isomorphic, as there could be different ways of simulating the same transformation. Also, there could be dynamics of the simulating device that do not correspond to dynamics in the thing being simulated. In general the group of dynamics possible in the simulator must *cover* the dynamics of the actual device.

using quantum interferometers.

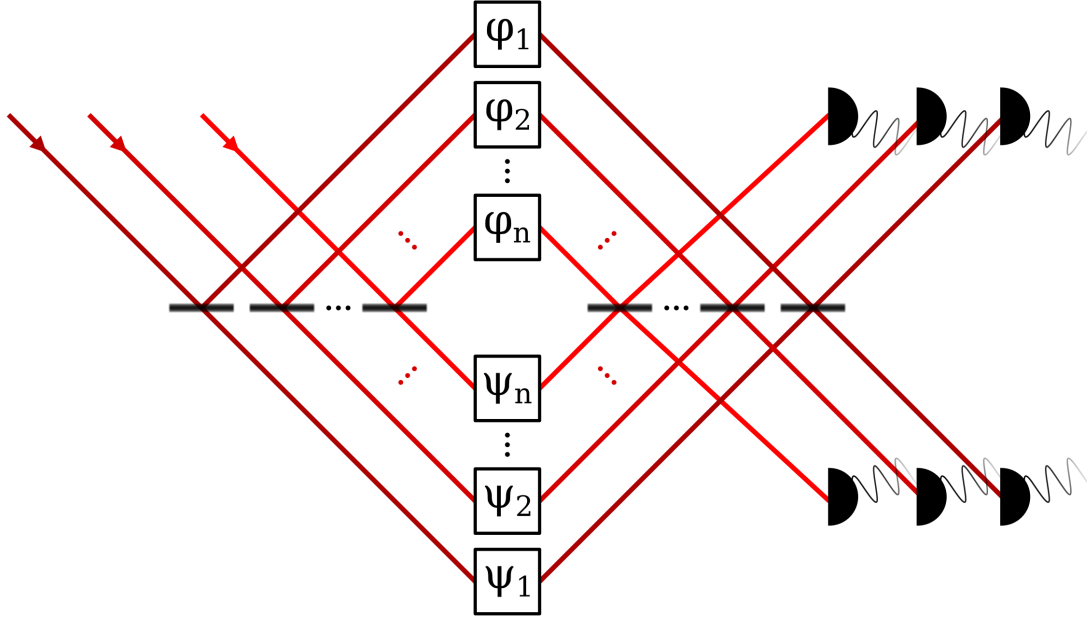


Figure 6.2: Quantum device exhibiting toroidal abelian interference. An element of the group $\mathcal{G}_\phi = \bigoplus^n \text{SO}(2)$ can be parametrised by a set of n angles $(\phi_1, \dots, \phi_n)$, and these angles can be used to configure the above quantum device.

Theorem 6.5. Any symmetrical interferometer that is compatible with relativity of simultaneity can be simulated using a number of quantum interferometers that scales linearly with the number of $\text{SO}(2)$ addends in the maximal torus of the phase group.

Proof. Recall from theorem 6.4 that the phase group of a symmetrical interferometer $\mathcal{G}_\phi$ that is compatible with relativity of simultaneity satisfies $\mathcal{G}_\phi \subseteq \bigoplus^n \text{SO}(2)$. We can build a quantum set up capable of implementing dynamics that cover this group by taking n two-branch complex quantum interferometers, arranged such that one branch of each goes to Alice and the other to Bob (see fig. 6.2).

Each element $T \in \bigoplus^n \text{SO}(2)$ may be parametrised by a set of angles $(\phi_1, \dots, \phi_n)$. In the simulation, these angles will correspond to putting a quantum phase plate ϕ_i in the arm associated with the interferometer i . (If we were to simulate a subgroup of the torus, then for some of the angles ϕ_i the agent is restricted to a finite subset of choices.) The choice of element T and its implementation can remain entirely local within one of the agents; once some element T has been chosen, Alice can implement phase plates entirely local to her branch, without having to communicate her choice to Bob.

Let the choice Alice made be $T_A \rightarrow (\phi_1, \dots, \phi_n)$ and the choice by Bob be $T_B \rightarrow (\psi_1, \dots, \psi_n)$. In this case, if the state of each interferometer in isolation is ρ_i , and the total state of the system ρ is written in the basis $\rho_1 \otimes \dots \otimes \rho_n$ (or a linear combination thereof for correlated or entangled states) the set of quantum interferometers will implement the following unitary operation:

$$\begin{aligned} U &= U_1 \otimes \dots \otimes U_n \\ &= \begin{pmatrix} e^{i\phi_1} & 0 \\ 0 & e^{i\psi_1} \end{pmatrix} \otimes \dots \otimes \begin{pmatrix} e^{i\phi_n} & 0 \\ 0 & e^{i\psi_n} \end{pmatrix}. \end{aligned} \quad (6.4)$$

This is a tensor product (rather than a direct sum), and it would be possible to input arbitrarily complex states (including those that are entangled between the different interferometers). For simplicity of argument, let us take the input to be a product state⁷. The output state ρ' will be $U\rho U^\dagger$ given by

$$\begin{aligned} \rho \rightarrow \rho' &= U\rho U^\dagger \\ &= U_1\rho_1 U_1^\dagger \otimes \dots \otimes U_n\rho_n U_n^\dagger. \end{aligned} \quad (6.5)$$

Define the operator X_i as the tensor product of many $\mathbb{1}_2$, replacing multiplicand i by the Pauli matrix σ_x :

$$X_i \equiv \mathbb{1}_2 \otimes \dots \otimes \mathbb{1}_2 \otimes \sigma_x \otimes \mathbb{1}_2 \otimes \mathbb{1}_2, \quad (6.6)$$

and equivalently Y_i with σ_y in position i .

Recall that for normalised states $\text{Tr}(\rho) = 1$, and for unitary operators $\text{Tr}(U\rho U^\dagger) = \text{Tr}(\rho)$. Also recall that $\text{Tr}(A \otimes B) = \text{Tr}(A) \text{Tr}(B)$. We can hence write:

$$\begin{aligned} \langle X_i \rangle' &= \text{Tr}(X_i U \rho U^\dagger) \\ &= \text{Tr}\left(U_1 \rho_1 U_1^\dagger \otimes \dots \otimes \sigma_x U_i \rho_i U_i^\dagger \otimes \dots \otimes U_n \rho_n U_n^\dagger\right) \\ &= \text{Tr}(U_i^\dagger \sigma_x U_i \rho_i). \end{aligned} \quad (6.7)$$

Likewise,

$$\langle Y_i \rangle' = \text{Tr}(U_i^\dagger \sigma_y U_i \rho_i). \quad (6.8)$$

We can explicitly write $U_i^\dagger \sigma_x U_i$ as

$$U_i^\dagger \sigma_x U_i = \begin{pmatrix} 0 & \exp(i(\psi_i - \phi_i)) \\ \exp(-i(\psi_i - \phi_i)) & 0 \end{pmatrix}, \quad (6.9)$$

⁷The following argument extends naturally to separable states: probabilistic combinations of product states. This extension introduces an extra sum into every equation where ρ appears, but does not ultimately not effect the final results in eq. (6.11) and eq. (6.12) because of the distributivity of matrix multiplication, and the linearity of trace.

and separate the real and imaginary parts such that

$$U_i^\dagger \sigma_x U_i = \cos(\psi_i - \phi_i) \sigma_x - \sin(\psi_i - \phi_i) \sigma_y. \quad (6.10)$$

Using the linearity of trace, we can thus express:

$$\begin{aligned} \langle X_i \rangle' &= \cos(\psi_i - \phi_i) \text{Tr}(\sigma_x \rho_i) - \sin(\psi_i - \phi_i) \text{Tr}(\sigma_y \rho_i) \\ &= \cos(\psi_i - \phi_i) \langle X_i \rangle - \sin(\psi_i - \phi_i) \langle Y_i \rangle, \end{aligned} \quad (6.11)$$

and similarly

$$\langle Y_i \rangle = \sin(\psi_i - \phi_i) \langle X_i \rangle + \cos(\psi_i - \phi_i) \langle Y_i \rangle. \quad (6.12)$$

Thus, we see the effect of the unitary U on each pair of measurements $\langle X_i \rangle$ and $\langle Y_i \rangle$ is to transform them according to an element of $\text{SO}(2)$, parametrised by angle $\psi_i - \phi_i$. These measurements can be made by applying the appropriate beam-splitter on interferometer i (e.g. a Hadamard gate to measure X_i). Thus, using a set of quantum interferometers, we can calculate exactly how two elements of $\bigoplus^n \text{SO}(2)$ interfere to transform a state.

□

We do not claim at this stage that this is the most efficient simulation: indeed, a quantum device of this dimension could be used for far more complicated dynamics those implemented in this simulation, and so it may be possible to create a smaller device that can implement the same dynamics.

For instance, one might recall from lemma 3.1 that the phase group of a maximal measurement on a $(n+1)$ -level quantum system is $(n+1)$ copies of $\text{U}(1)$, $\text{U}(1) \oplus \dots \oplus \text{U}(1)$: which is itself isomorphic to a torus $\text{SO}(2) \oplus \dots \oplus \text{SO}(2)$. Thus, we can use a single $(n+1)$ -level quantum device to simulate $\bigoplus^n \text{SO}(2)$ (the extra dimension is needed because the global phase is not measurable), and implement each difference in angle $\psi_i - \phi_i$ via a phase plate on branch $i + 1$ (leaving branch 1 as a reference). This would execute the dynamics of the torus, and the output statistics can be measured by placing an appropriate beam-splitter between branch 1 with branch i .

This is a less satisfying simulation than the set-up presented above, because the system no longer naturally divides into two separate agents who can independently choose the two sets of interfering angles. Here, the simulation is more forced, in that each $\psi_i - \phi_i$ is calculated locally (at one branch) beforehand, and the result programmed manually into the operation of the device. The experimenter reading the output no longer really witnesses interference between the two choices of phase group element, but is rather just receiving direct communications from each branch individually.

6.3.1 Remark: Random access Deutsch's algorithm

The phase group $\mathcal{G}_{ab} \subseteq \bigoplus^n \text{SO}(2)$ is somewhat similar to the qubit phase group (for $n = 1$ it is exactly the phase group of a single qubit). However, an $n > 1$ device can perform an algorithm that a single qubit could not do: a random access variant of Deutsch's algorithm [21, 22].

Let us write the state-space of systems acted on with this toroidal phase group as

$$\mathbf{s} = (\langle Z \rangle, \langle X_1 \rangle, \langle Y_1 \rangle, \dots, \langle X_n \rangle, \langle Y_n \rangle)^T, \quad (6.13)$$

where $\langle Z \rangle$ is the position measurement (fixed by the phase group), and each pair $\langle X_i \rangle$ and $\langle Y_i \rangle$ is acted upon by one of the $\text{SO}(2)$. Agents Alice and Bob, by restricting the phase group choices to the subgroup $Z_2 \oplus Z_2 \oplus \dots \oplus Z_2$, can encode a string of n bits of their choosing on each arm of the interferometer. An experimentalist, through cunning choice of input state or output measurement (i.e. selection of beam-splitter; see chapter 4) can perform an experiment that identifies in one shot whether Alice and Bob have chosen the same value for one bit in the string. For instance, to find bit i , the experimentalist would prepare a (post-beam-splitter) input state $\langle Z \rangle = 0$, $\langle X_i \rangle$ to $+1$, and $\langle Y_i \rangle = 0$ (all other measurements are not important, though their outcomes may be constrained by restrictions on which states are allowed in the theory). The experiment would then transform $\langle X_i \rangle$ to $+1$ if the bits are the same, and -1 if they differ.

This is a *random-access* algorithm, as it works to determine the answer for one of n bits encoded, but the encoding agents Alice and Bob do not need to know in advance which bit the experimentalist will examine. Within the simulation described in theorem 6.5, the performance of the random access Deutsch algorithm (§ 6.3.1) becomes obvious: one makes n copies of single-qubit implementations of Deutsch's algorithm, and discards the output of all but the one bit of interest. On the other hand, in the $(n+1)$ -level quantum simulation described above, the experiment becomes totally trivial, as in order to program the correct behaviour (i.e. select the correct phase plate), at each branch one has to already know whether a particular bit is the same or different for the two choices of bit-strings.

6.4 n -balls and the dimensionality of quantum theory

Let us now consider the results of § 6.2 as applied to a more specific class of theories: those whose state-space is given by an n -ball. An n -ball is a natural generalisation of the qubit state-space into an arbitrary number of measurements, defined as the set of

real vectors of the form $\mathbf{X} = (X_1, \dots, X_n)^T \in \mathbb{R}^n$ subject to the constraint

$$\sum_{i=1}^n |X_i|^2 \leq 1. \quad (6.14)$$

When drawn in Euclidean space, this yields a circle for $n = 2$, and a sphere⁸ for $n = 3$, both of unit radius centred around the origin.

Such shapes appear in the GPT framework as the state-space of expectation values associated with a set of n binary measurements (with outcomes ± 1), subject to a restriction on making simultaneous predictions of the outcomes. Indeed for measurements M_i , we can use the mapping $P(M_i = +1) = \frac{1}{2}(1 + X_i)$ and $P(M_i = -1) = \frac{1}{2}(1 - X_i)$ to convert the ball vector $\mathbf{X}$ into a state written in terms of probabilities.

The automorphisms of the n -ball are given by the orthogonal group of dimension n , written $O(n)$. This consists of rotations (in up to n dimensions) and reflections. If we only consider the portion of the group connected to the identity, then we have the special orthogonal group $SO(n)$ that does not admit reflections. The phase group associated with any maximal measurement on an n -ball fixes an axis, and so is given by $O(n-1)$ if reflections are allowed, or $SO(n-1)$ otherwise.

Ball state spaces arise naturally in the axiomatisation of quantum mechanics (e.g. [9, 14, 17]). They correspond to set-ups where exactly two possibilities can be perfectly distinguished, but there is an infinite number of possible such pairings (for $n > 1$) which can be continuously deformed into each other. In particular, by selecting the $n = 3$ ball, one arrives at the standard complex-valued quantum theory qubit.

It is also known that in general (for any dimension) ball state spaces correspond to two-level systems described by Jordan algebras [93, 94, 111]. Alongside $n = 3$ for standard complex quantum theory, we also have $n = 2$ corresponding to real-valued quantum theory, $n = 5$ for quaternionic⁹ quantum theory [112, 113], and $n = 9$ for an octonionic two-level system.

We shall see that there is a strong limit on how high the dimension of ball can be for all its valid phase dynamics to be compatible with relativity of simultaneity.

Lemma 6.6. *When the only constraint on locally applicable phase transformations is*

⁸ We must take care in this section to distinguish between ‘sphere’ and ‘ball’. An n -ball is the compact convex set defined in eq. (6.14); an $(n-1)$ -sphere is the non-convex surface of that set where $\sum_{i=1}^n |X_i|^2 = 1$ and is labelled with one fewer dimension, reflecting the degrees of freedom in moving around such a surface. In this terminology, what would be commonly called the Bloch sphere should in fact be referred to as the Bloch (3-)ball.

⁹ The quaternions $\mathbb{H}$ can be thought of as an extension to the field of complex numbers such that there are three types of imaginary number. A quaternion $H \in \mathbb{H}$ may be expressed as $H = a + ib + jc + kd$, where $a, b, c, d \in \mathbb{R}$ and $ii = jj = kk = ijk = -1$. For a lucid discussion in a quantum context, see Graydon [112].

that they must be compatible with branch locality, the highest dimension n -ball with dynamics compatible with relativity of simultaneity is $n = 2$ if reflections are allowed, or $n = 3$ otherwise.

Proof. States described by an n -ball satisfy a quantum-like uncertainty relation (definition 5.4) for any dimension. When $X_i = \pm 1$ corresponding to M_i being fully determined such that either $P(M_i = +1) = 1$ or $P(M_i = -1)$, then from eq. (6.14) all other $X_j = 0$ for $j \neq i$, implying that all other measurements are uniformly random such that $P(M_j = +1) = P(M_j = -1) = \frac{1}{2}$.

From this relation and lemma 5.8, we see that hyper-spheres are fully conditionally restricted with respect to any measurement. Moreover, when considered as an interferometer (such that one of the measurements encodes the position), then from corollary 5.7 of theorem 5.6, all of the phase dynamics are consistent with branch locality.

However, from corollary 6.3 the phase group must also be abelian. $O(1)$ is the largest abelian orthogonal group, corresponding to the $\{1, \text{flip}\}$ phase group of a circular state space dimension $n = 2$. $SO(2)$ is the largest abelian special orthogonal group, correspond to the rotations of a sphere around a fixed axis in dimension $n = 3$. $\square$

6.4.1 Subgroups of the orthogonal group

We can in fact go a step further, and also consider if there is some abelian subgroup of $O(n)$ (as we have been implicitly doing for $SO(n) \subset O(n)$) that would allow for state spaces of larger dimension. For instance, the largest abelian subgroup of $O(2n)$ and $O(2n + 1)$ will be bounded by the maximal torus $\bigoplus_{i=1}^n SO(2)$ (see e.g. [110]), and the discussion in theorem 6.4). Inspired by the axiomatic approaches of, e.g., [9] or [14], we might be able to rule out a lot of these theories by postulating other properties that the dynamics of the object should satisfy.

Let us take as our first such restriction that the full group of dynamics (i.e. before we freeze an axis to find the phase group) must be connected. This could follow from time-continuity of dynamics [9], where for every transformation, it should be possible (especially in the limit of small times) to find another that corresponds to doing the same thing for a fraction of the time, thus inducing a fraction of the effect. This is only possible when every element in the group is connected to the identity element (corresponding to implementing an action for no time at all). As such, this restriction rules out all the finite groups (which are naturally discrete rather than continuous), as well as groups such as $O(n)$ that contain reflections which cannot be subdivided.

Next, we consider the requirement that for every pair of pure states, there should be at least one transformation that maps one pure state to the other. This is a property

known as transitivity on the sphere (surface of the ball) [81, 114]. With this restriction, we can rule out a lot of theories.

For orthogonal groups, one can find subgroups of the form $O(n_1) \oplus \cdots \oplus O(n_m) \subseteq O(n_1 + \cdots + n_m)$. One might attempt to form an abelian phase group by locking out one axis only, from a group of the form $G \oplus \text{SO}(2) \oplus \cdots \oplus \text{SO}(2)$, where the phase condition must induce the subgroup $\text{SO}(2)$ from G (e.g. G could be $\text{SO}(3)$). However, it is easy to see that groups with this direct-product structure cannot be transitive on the surface of the sphere.

Lemma 6.7. *Any group formed from the direct product of two or more groups, where one group is orthogonal ($O(n_1) \oplus G$) cannot be transitive on the surface of an n -ball.*

Proof. Consider an n -ball with $1 \leq n_1 < n$. One can find two different valid states $\mathbf{X}$ and $\mathbf{Y}$ on the surface:

$$\begin{aligned}\mathbf{X} &= (a_1, \dots, a_{n_1}, \dots)^T, \\ \mathbf{Y} &= (b_1, \dots, b_{n_1}, \dots)^T,\end{aligned}\tag{6.15}$$

such that $a_1^2 + \cdots + a_{n_1}^2 \neq b_1^2 + \cdots + b_{n_1}^2$.

From the direct-sum nature of the group, the first n_1 elements of the vector are only acted on by $O(n_1)$. However, the orthogonal group preserves the 2-norm of vectors, such that no element of n_1 acting taking the vector $(a_1, \dots, a_{n_1})^T$ to $(a'_1, \dots, a'_{n_1})^T$ will ever result in $a_1'^2 + \cdots + a_{n_1}'^2 \neq a_1^2 + \cdots + a_{n_1}^2$. Hence, there is no operation in the group that will transform $(a_1, \dots, a_{n_1})^T$ to $(b_1, \dots, b_{n_1})^T$. Thus $O(n_1) \oplus G$ cannot be transitive on the surface of an n -ball. $\square$

Let us now consider more sophisticated subgroups of $O(n)$. It is known that $U(n) \subset \text{SO}(2n)$, which can be seen by embedding an $n \times n$ complex unitary matrix U into a $2n \times 2n$ real matrix

$$T_U = \begin{pmatrix} \Re(U) & \Im(U) \\ -\Im(U) & \Re(U) \end{pmatrix}\tag{6.16}$$

and verifying (such as by direct multiplication) that T_U is a representation of $U(n)$ and that T_U are orthogonal matrices. Indeed, following this prescription gives us the equivalence between $U(1) = \{e^{i\phi}\}$ and $T_{U_1} \in \text{SO}(2)$ such that

$$T_{U_1} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix},\tag{6.17}$$

which we have been using implicitly in the treatment of the qubit phase group throughout this thesis.

Matrices of this form are in fact transitive on the sphere [81] but only $U(2)$ has a phase group that is non-trivial and abelian.

Lemma 6.8. *The only unitary subgroup of $O(n)$ with a non-trivial abelian phase-group corresponds to $U(2) \subset SO(4)$ in a four-dimensional ball. This phase group is isomorphic to $SO(2)$.*

Proof. Take the first element of the state vector to correspond to the measurement whose phase group we want to find. In order to preserve this statistic for all states, the top row of T_U must be $(1, 0 \dots 0)$. This implies that the top row of the unitary matrix U must be $(1, 0 \dots 0)$. As U is unitary, this also means that the left-most column of U is $(1, 0 \dots 0)^T$, and hence the phase group is given by the block-diagonal unitary matrix $\mathbb{1} \oplus U(n-1)$, which is only abelian for $n \leq 2$.

- For $U(1)$, the phase group is trivial (fixing the top row fixes every element in the matrix, such that the phase group is $\{\mathbb{1}_2\}$).
- For $U(2)$, the phase group is given by $U = \text{diag}(1, e^{i\phi})$ which in the four-dimensional real representation may be written

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \phi & 0 & \sin \phi \\ 0 & 0 & 1 & 0 \\ 0 & -\sin \phi & 0 & \cos \phi \end{pmatrix}, \quad (6.18)$$

which is a representation of $SO(2)$ acting on the second and fourth elements.

Alternative formulations of the representation T_U in eq. (6.16) would allow us to change which two measurements are preserved. $\square$

Similarly, there are ways of writing the hyperunitary (or symplectic) group $\text{Sp}(n)$ as a subgroup of real orthogonal matrices $SO(4n)$. Hyperunitaries are the natural generalisation of unitary operations such that they preserve the inner product of a quaternionic Hilbert space [115]. In the same way that for an orthogonal matrix $O^T O = \mathbb{1}$, and for a (complex) unitary matrix $U^\dagger U = \mathbb{1}$, elements in the hyperunitary group satisfy $H^\dagger H$ where $[\cdot]^\dagger$ is the quaternionic equivalent of hermitian conjugation (transpose the matrix and conjugate each element such that $H = a + ib + jc + kd$ maps to $\bar{H} = a - ib - jc - kd$).

The groups $\text{Sp}(n)$ are connected, and transitive on the surface of the sphere of dimension $4n$ [81]. However, we can see that for all n , they are either trivial or non-abelian.

Lemma 6.9. *There are no hyperunitary subgroups $\text{Sp}(n)$ of $O(4n)$ with a non-trivial abelian phase-group.*

Proof. For some $n \times n$ hyperunitary matrix H , we can take the real and three imaginary parts to form four real matrices of the same dimension: $R = \Re(H)$, $I = \Im(H)$, $J = \mathbb{J}\Im(H)$ and $K = \mathbb{K}\Im(H)$ (where the operations $\Im$, $\mathbb{J}\Im$ and $\mathbb{K}\Im$ take the components associated with i , j and k respectively).

Quaternionic matrices can be embedded into a real matrix in such a way that they preserve the correct form of quaternionic multiplication (verifiable by direct multiplication):

$$H \in \mathbb{H}^{n \times n} \leftrightarrow T_H = \begin{pmatrix} \Re H & \Im H & \mathbb{J}\Im H & \mathbb{K}\Im H \\ -\Im H & \Re H & -\mathbb{K}\Im H & \mathbb{J}\Im H \\ -\mathbb{J}\Im H & \mathbb{K}\Im H & \Re H & -\Im H \\ -\mathbb{K}\Im H & -\mathbb{J}\Im H & \Im H & \Re H \end{pmatrix}. \quad (6.19)$$

When $H \in \text{Sp}(n)$, then $T_H \in \text{SO}(4n)$, which may be seen by noting that the quaternionic conjugate $H^\dagger$ gives the matrix $T_{H^\dagger} = T_H^T$.

As with lemma 6.8, finding the phase group which leaves one axis invariant will require the top row (and hence left-column) to be $(1, 0, \dots, 0)$, thus inducing the phase group $\mathbb{1}_1 \oplus \text{Sp}(n-1)$. When $n = 1$, this is trivial, and when $n \geq 2$, this is $\text{Sp}(n-1)$, which is non-abelian. $\square$

The orthogonal, unitary and hyperunitary groups are the only groups that are transitive on hyper-spheres of arbitrary dimension. However, there are also a few special cases that exist for particular dimensions. An exhaustive list is given by Kerin and Wraith [81], as well as in chapter 2 in volume II of [39]. These exceptional cases are:

- $n = 7$: $G_2 \subset \text{O}(7)$ with phase group $\text{SU}(3)$,
- $n = 8$: $\text{Spin}(7) \subset \text{O}(8)$ with phase group G_2 ,
- $n = 16$: $\text{Spin}(9) \subset \text{O}(16)$ with phase group $\text{Spin}(7)$,

all three of which have non-abelian phase groups.

6.4.2 Interpretation of the four-dimensional case and synopsis

Out of all the transitive connected subgroups of $\text{O}(d)$, we have only found two examples with non-trivial abelian phase groups: $\text{SO}(3)$ in $d = 3$ and $\text{U}(2)$ in $d = 4$.

At first, it appears odd that we have found a 4-dimensional ball in addition to the usual Bloch ball. In the context of the two-level interferometer, these two examples are identical in their dynamics: even with access to the full dynamics of the system, every unitary operation $U \in \text{U}(2)$ corresponds to a rotation $R_U \in \text{SO}(3)$.

When we treat these groups more rigorously, this is not quite the full story. It is the case that the action of every unitary matrix $U \in \text{U}(2)$ on a density matrix $\rho \rightarrow U\rho U^\dagger$ can be represented by a rotation $R_U \in \text{SO}(3)$ on the Bloch sphere. However, despite this $\text{U}(2)$ and $\text{SO}(3)$ are not isomorphic. In fact, for any real ϕ , $U' = e^{i\phi}U$ will have the same effect on ρ as U , and thus the mapping from $\text{U}(2) \rightarrow \text{SO}(3)$ is many-to-one. As this extra freedom is in itself a one-dimensional unitary $\text{U}(1)$, the relationship is expressed as the quotient $\text{U}(2)/\text{U}(1) \cong \text{SO}(3)$.

Thus, we see that the four-dimensional representation of the $\text{U}(2)$ group does contain more information than the $\text{SO}(3)$ representation: it also encodes information about the global phase. By placing only dynamical restrictions on the theories that could describe a two-level interferometer, in both cases we have arrived at the $\text{U}(2)$ dynamic corresponding to complex quantum theories. That this results in two possibilities differing by one dimension follows from whether we encode information about the global phase. On the other hand, the phase groups of both cases, $\text{U}(1)$ and $\text{SO}(2)$, are truly isomorphic and so the behaviour of interference patterns in a Mach-Zehnder interferometer is exactly the same.

We summarise the results of this section with the following theorem:

Theorem 6.10. *Suppose we have a two-branch interferometer described by an n -ball, and make the following demands:*

- i. the phase group is non-trivial,*
- ii. the phase group is compatible with relativity of simultaneity (i.e. abelian),*
- iii. all elements of the phase group can be applied on either branch (as is compatible with branch locality),*
- iv. the full dynamics can map any pure state to any other (i.e. are transitive on the surface of the sphere),*
- v. the full dynamics respect time-continuity (i.e. are connected).*

The only two allowed groups of transformations that satisfy all of the above are for:

- A. $n = 3$ with $\text{SO}(3)$ dynamics, and $\text{SO}(2)$ phase group: corresponding to qubit density matrices represented by the Bloch ball in complex quantum mechanics.*
- B. $n = 4$ with $\text{U}(2)$ dynamics, and $\text{SO}(2)$ phase group: where the surface corresponds to pure states in complex quantum mechanics including global phase.*

6.5 Conclusion and outlook

In this chapter we considered how the operational description of the state of an interferometer should behave in the context of special relativity. We argued that since probabilities reflect objective reality, inner products of the form $\langle e | T | s \rangle$ should be Lorentz invariant. Moreover, when s represents the state of a Mach–Zehnder interferometer, we showed that the transformations on the state must also be the same in every reference frame.

We concluded that for this result to be consistent with relativity of simultaneity, the phase operations local to different branches of an interferometer must commute with each other. This places a strong restriction on the state-spaces that we can admit as descriptions of the Mach–Zehnder interferometer, and in the specific context of n -balls lead us to two possibilities: $n = 3$ corresponding to standard quantum theory, and $n = 4$ which has exactly the same dynamics as quantum theory in this context, but might allow for different measurements.

The exact interpretation of the $n = 4$ case (which also appeared recently in [116]) is an interesting question still to be answered. Moreover, there has been recent experimental interest in the testing for interference patterns beyond those that could be explained by quantum theory [117–119] (in particular [119] is motivated by a proposal by Peres [120], which distinguishes between complex and quaternionic quantum theory). However, the results in this chapter show us that one must be careful in designing the experiment; if it involves spatially disjoint branches (as most optical set-ups will), the rules of relativity could prevent us from observing post-quantum behaviour.

Chapter 7

Post-quantum particle exchange

In chapter 3, we extended the notation of phase from quantum theory into the GPT framework, and through-out chapters 4 to 6 we explored its application in the context of interferometry. In this chapter, we apply phase to a slightly different quantum phenomena: that of particle statistics.

In three-dimensional free space, quantum particles come in two types: bosons and fermions. The physical behaviour of bosons and fermions can be significantly different, especially when one considers an *indistinguishable* set of them. Fermions obey the Pauli exclusion principle, with wide applications ranging from the existence of neutron stars when applied to neutrons, to the entire field of semi-conductor physics when applied to electrons (see e.g. [121]). On the other hand, bosons display bunching behaviour, and at low temperatures can form unusual states of matter known as Bose-Einstein condensates (see e.g. [122–124]).

Ultimately, these divergent behaviours follow from a subtle but crucial difference in how the phase of the wave-function of the system changes when one swaps one indistinguishable particle with another. As this critical difference between bosons and fermions lies in a *phase*, it is natural for us to consider if any of the exotic phase dynamics that have appeared for theories in the GPT framework might themselves be associated with post-quantum particles.

In this chapter¹, we shall begin with a mathematical review of exchanging subsystems within quantum theory (§ 7.1), and how this corresponds to different particle types. We shall then define indistinguishability in the GPT framework, and provide a physical mechanism for detecting different particle types in arbitrary theories (§ 7.2). We shall conclude by considering a few examples of what type of post-quantum particles we might observe (§ 7.3).

7.1 Exchanging subsystems in quantum theory

First let us explore the concept of swapping subsystems in quantum theory. We begin with the simple example of product states, in order to clearly define the swap

¹Some of the concepts in this chapter were presented in [4].

operation, before moving on to arbitrary quantum states in the next subsection.

Suppose there is a system, which admits description by some pure state $|\Psi\rangle$, and another another copy of the system, described by an alternative pure state $|\Phi\rangle$. In quantum theory, we can write down the joint state of the two systems when they are uncorrelated by using a tensor product $|\Psi\rangle|\Phi\rangle \equiv |\Psi\rangle \otimes |\Phi\rangle$. The tensor product between two complex Hilbert spaces is well defined, and so measurements made on system 1 have the form $M_1 = M \otimes \mathbb{1}$, and those on system 2 are $M_2 = \mathbb{1} \otimes M$, such that when the states are normalised, $P(M \otimes \mathbb{1}|\Psi\Phi) = \langle\Psi|M|\Psi\rangle$ and $P(\mathbb{1} \otimes M|\Psi\Phi) = \langle\Phi|M|\Phi\rangle$.

Consider now an operation $T(\cdot)$ that swaps these two systems around, such that M on the first system in the new state gives the same results that M on the second system of the old did (and vice versa) thus satisfying:

$$\begin{aligned} P(M \otimes \mathbb{1}|T(\Psi\Phi)) &= P(\mathbb{1} \otimes M|\Psi\Phi) \quad \forall M \\ P(\mathbb{1} \otimes M|T(\Psi\Phi)) &= P(M \otimes \mathbb{1}|\Psi\Phi) \quad \forall M. \end{aligned} \quad (7.1)$$

More strictly, we even require that correlation of measurement M on system 1 with measurement N on system 2 beforehand is transformed into the correlation of N on system 1 and M on system 2 afterwards:

$$P(N \otimes M|T(\Psi\Phi)) = P(M \otimes N|\Psi\Phi) \quad \forall M, N. \quad (7.2)$$

The duality of states and measurements in quantum theory tells us that the most general operation satisfying the above must exchange the ordering of $|\Psi\rangle$ and $|\Phi\rangle$ in the tensor product, and thus its action on product states is to take

$$T : |\Psi\rangle \otimes |\Phi\rangle \rightarrow e^{i\phi} |\Phi\rangle \otimes |\Psi\rangle. \quad (7.3)$$

This extra freedom ϕ is a phase term that appears because there are no quantum measurements can detect global phase.

Suppose we go a step further such that not only do the two systems share the same state-space, but are both prepared in the same pure state (i.e. $|\Psi\rangle = |\Phi\rangle$). In this case, the swap operation should cause no detectable change in statistics, and so the product state evolves as $|\Psi\rangle|\Phi\rangle \rightarrow e^{i\phi} |\Psi\rangle|\Phi\rangle$.

When the two systems are measured in isolation, the term in ϕ is an unobservable global phase, and there is no way of telling if a swap was performed or not. However, this changes if we introduce a third system $|\xi\rangle$ as a reference. Let $|\xi\rangle$ be a qubit control for the swap operation, such that if $|\xi\rangle = |0\rangle$, then no swap is performed, and if $|\xi\rangle = |1\rangle$, the swap is executed. In this sense, $|\xi\rangle$ is the *control bit* for a *controlled-swap* operation (also known as a *Fredkin gate* when the systems being exchanged are qubits and $\phi = 0$).

Quantum theory allows $|\xi\rangle$ to be in a superposition, such that the controlled-swap operation puts $|\Phi\rangle$ and $|\Psi\rangle$ into a superposition of being swapped and not swapped. Physically, this could correspond to putting the two particles into a superposition of spatial locations, and then applying some manipulation to swap it in only one of these locations, and leaving it unchanged in the other (see fig. 7.1).

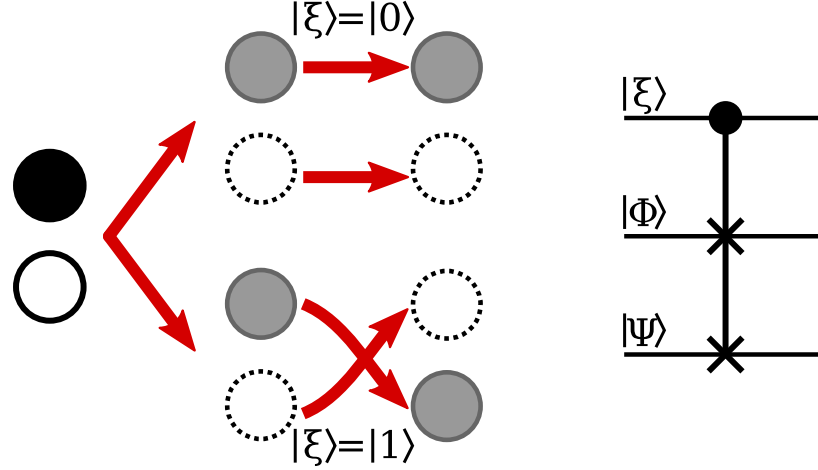


Figure 7.1: Quantum controlled-swap device. Two particles (drawn black and white) enter a superposition of locations. In the top location ($|\xi\rangle = |0\rangle$), no action is taken; whereas in the bottom location ($|\xi\rangle = |1\rangle$) the particles are exchanged. Such a device implements the quantum controlled-swap gate, illustrated by the circuit component on the RHS, where the spatial location freedom $|\xi\rangle$ is used as the control bit.

Writing $|\xi\rangle = \alpha|0\rangle + \beta|1\rangle$ (for $\alpha, \beta \in \mathbb{C}$), the action of this controlled-swap is to take:

$$|\xi\rangle|\Phi\rangle|\Psi\rangle \xrightarrow{c\text{-swap}} \alpha|0\rangle|\Phi\rangle|\Psi\rangle + e^{i\phi}\beta|1\rangle|\Psi\rangle|\Phi\rangle. \quad (7.4)$$

Suppose there are now no distinguishing features between the two systems being exchanged such that $|\Psi\rangle|\Phi\rangle = |\Phi\rangle|\Psi\rangle$. In this case, eq. (7.4) simplifies to

$$|\xi\rangle|\Psi\rangle|\Phi\rangle \xrightarrow{c\text{-swap}} (\alpha|0\rangle + e^{i\phi}\beta|1\rangle) |\Psi\rangle|\Phi\rangle, \quad (7.5)$$

and we see that the global phase of the swap in isolation now causes a measurable phase transformation $\text{diag}(1, e^{i\phi})$ on the control bit.

7.1.1 Indistinguishable quantum states

We have identified a family of controlled-swap operations, with a $U(1)$ freedom in possible phase transformations effected on the control bit. However, the choice of

control-swap gate is not the only method by which different phases may be induced on the control bit.

The above section discussed product states, but let us now consider pure quantum states that might also have correlations. Taking two arbitrary (pure) states of a single system as $|\chi\rangle$ and $|\xi\rangle$. Ignoring normalisation, for a bi-partite system (with subsystems A and B), we could prepare a symmetrised state

$$|\Psi\rangle_S = |\chi\rangle|\xi\rangle + |\xi\rangle|\chi\rangle, \quad (7.6)$$

or anti-symmetrised state:

$$|\Psi\rangle_{AS} = |\chi\rangle|\xi\rangle - |\xi\rangle|\chi\rangle. \quad (7.7)$$

No local measurement made at either A or B will offer different statistics for either of these wave-functions. Moreover, even knowing the correlations between measurements made on A and B will not reveal a preferred labelling for A and B : the projection onto any pure product state $|m\rangle \otimes |n\rangle$ (where $|m\rangle$ and $|n\rangle$ may be in the same or differing bases) has the same probability as a projection onto $|n\rangle \otimes |m\rangle$. For these wave-functions, exchanging every instance of χ with ξ only results in a change of the wave-function by global phase (by 0 or π respectively), the effect of which is unmeasurable for a bipartite system in isolation.

We remark that a naïve attempt to form an ‘anyonic’ (e.g. [125, 126]) symmetrisation such as

$$|\Psi\rangle_{\text{any}} = |\chi\rangle|\xi\rangle + e^{-i\alpha}|\xi\rangle|\chi\rangle \quad (7.8)$$

does not satisfy the above. This may be seen in the following two-qubit counter-example: take $|\chi\rangle = |0\rangle$ and $|\xi\rangle = |+\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle)$, and consider the projection onto $|m\rangle|n\rangle$ vs. $|n\rangle|m\rangle$ where $|m\rangle = 1/\sqrt{2}(|0\rangle + i|1\rangle)$ and $|n\rangle = |0\rangle$. In this case, $|\langle mn|\Psi\rangle_{\text{any}}|^2 = \frac{1}{8}(3 + 2(\cos \alpha + \sin \alpha))$ whereas $|\langle nm|\Psi\rangle_{\text{any}}|^2 = \frac{1}{8}(3 + 2(\cos \alpha - \sin \alpha))$. These are only equal for $\alpha = 0$ or π . For all other α , these statistics will differ, allowing us to compare these statistics and establish which side is A and which is B . Moreover, an attempt to exchange χ with ξ results in the state $|\Psi'\rangle_{\text{any}} = e^{-i\alpha}(|\chi\rangle|\xi\rangle + e^{+i\alpha}|\xi\rangle|\chi\rangle)$ where the relative phase between the two orderings is taken to its complex conjugate, and hence can correspond to a measurably different state except when $e^{i\alpha} = 1$ or -1 .

In quantum theory, the *type* (or *species*) of a particle is assigned a priori, perhaps by some other rule (the spin statistics theorem [127, 128] states that particles with integer spin, such as photons, are *bosonic*, and those with half-integer spin, such as electrons, are *fermionic*). Once the type has been determined, this places a rule about which wave-functions are allowed as a valid description of a systems of identical particles. A pair of identical bosons must be in the symmetric form $|\Psi\rangle_S$, a pair of identical fermions in the form $|\Psi\rangle_{AS}$.

Identical particles of the same species thus have a wave-function that obeys strict

symmetry rules, and this lies behind some of the most drastic quantum effects, such as the Pauli exclusion principle (i.e. identical fermions cannot share the same state as otherwise $|\Psi\rangle_{AS} = 0$), and form the basis of quantum particle statistics (i.e. the Fermi-Dirac and Bose-Einstein distributions).

Let us now consider what happens when we perform a controlled-swap of the subsystems of an (anti-)symmetrised wave-function of the form $|\Psi\rangle = |\chi\rangle|\xi\rangle + e^{-i\alpha}|\xi\rangle|\chi\rangle$ for $\alpha \in \{0, \pi\}$. This has the effect:

$$|\xi\rangle|\Psi\rangle \xrightarrow{c\text{-}swap} (\alpha|0\rangle + e^{i(\phi+\alpha)}\beta|1\rangle) |\Psi\rangle. \quad (7.9)$$

The phase freedom in how the state has been symmetrised combines with the phase freedom in the swap operation, to shift the phase of the control bit by $\phi+\alpha$. If ϕ remains as a constant for any physical swap operation, then the wave-functions with different α will result in different final states of the control bit. Hence, in this experiment, we may identify a different type of symmetrisation for each element in the quantum phase group associated with the control bit.

For example (setting $\phi = 0$), and taking the control bit to initially be in state $|+\rangle$, the symmetric input results in

$$|+\rangle|\Psi\rangle_S = |0\rangle|\Psi\rangle_S + |1\rangle|\Psi\rangle_S \xrightarrow{c\text{-}swap} |0\rangle|\Psi\rangle_S + |1\rangle|\Psi\rangle_S = |+\rangle|\Psi\rangle_S, \quad (7.10)$$

and the control bit evolves as $|+\rangle \rightarrow |+\rangle$. For the anti-symmetric case

$$|+\rangle|\Psi\rangle_{AS} = |0\rangle|\Psi\rangle_{AS} + |1\rangle|\Psi\rangle_{AS} \xrightarrow{c\text{-}swap} |0\rangle|\Psi\rangle_{AS} - |1\rangle|\Psi\rangle_{AS} = |-\rangle|\Psi\rangle_{AS}, \quad (7.11)$$

and the control bit evolves as $|+\rangle \rightarrow |-\rangle$. If the particle was guaranteed to be either fermionic or bosonic, then a X measurement on the final state of the control bit would deterministically reveal which of these two alternatives is the case in a single measurement, without the need to directly measure the symmetrised wave-function.

7.2 Swapping and indistinguishability in post-quantum theories

Although a lot of the above discussion was deeply steeped in the structure of complex Hilbert spaces, we will now consider how much of the above quantum reasoning can be extended to arbitrary theories within the GPT framework. We shall first probe what it means for the subsystems of a bipartite system to be indistinguishable.

Consider a bi-partite system AB (see § 2.6) where the set of possible (reduced) states in A are isomorphic to those in B , and likewise the possible measurement effects

for A are isomorphic to the set for B . In direct analogy with the quantum case eq. (7.2), let us define a transformation T_{swap} that acts on AB :

Definition 7.1: Swap operation

A **swap operation** T_{swap} is a transformation that satisfies

$$(\mathbf{e}_A \otimes \mathbf{e}_B) \cdot \mathbf{s}_{AB} = (\mathbf{e}_B \otimes \mathbf{e}_A) T_{\text{swap}} \mathbf{s}_{AB}, \quad (7.12)$$

for all $\mathbf{e}_A$ and $\mathbf{e}_B$, where $\{\mathbf{e}_A\} = \{\mathbf{e}_B\}$ are the measurement effects possible on either subsystem.

When the state-space of AB is locally tomographic, measuring the set of all $\mathbf{e}_A \otimes \mathbf{e}_B$ is sufficient to completely specify the state $\mathbf{s}_{AB}$ (which might also include correlations between the two subsystems). For locally tomographic theories, definition 7.1 will thus be sufficient to define an operationally unique swap operation on any state $\mathbf{s}_{AB}$ of the joint system². (This does not contradict the $U(1)$ global phase freedom in quantum theory as this degree of freedom is non-operational!)

Let us now consider an operational definition for subsystems to be indistinguishable in general (regardless of whether or not the theory is locally tomographic).

Definition 7.2: Subsystems indistinguishable under exchange

Two subsystems A and B are defined to be **indistinguishable under exchange**, if they are in a state such that exchanging them has no observable effect. That is, for a swap operation T_{swap} (satisfying eq. (7.12)), the state $\mathbf{s}_{AB}$ represents indistinguishable A and B if

$$T_{\text{swap}} \mathbf{s}_{AB} = \mathbf{s}_{AB}. \quad (7.13)$$

We remark that symmetrised and anti-symmetrised quantum states will satisfy this definition for indistinguishability. This definition will also admit the possibility of ‘anyonic’ states of quantum theory where the swap operation induces an arbitrary, but unmeasurable, global phase on the joint system [125, 126].

7.2.1 Locally tomographic theories, and local distinguishability

For product states in a locally tomographic theory, where $\mathbf{s}_{AB} = \mathbf{s}_A \otimes \mathbf{s}_B$, then $\mathbf{s}_A = \mathbf{s}_B \equiv \mathbf{s}$, is sufficient condition for indistinguishability. This is because $(\mathbf{e}_A \otimes \mathbf{e}_B) \cdot \mathbf{s}_{AB} = (\mathbf{e}_A \cdot \mathbf{s})(\mathbf{e}_B \cdot \mathbf{s}) = (\mathbf{e}_B \cdot \mathbf{s})(\mathbf{e}_A \cdot \mathbf{s}) = (\mathbf{e}_B \otimes \mathbf{e}_A) \cdot \mathbf{s}_{AB}$.

²For theories that are not locally tomographic, it is also necessary to specify what happens to the “global” freedoms that are neither local statistics or correlations.

Even in locally tomographic theories, local distinguishability (that $s_A = s_B$) is not sufficient to imply that a joint state s_{AB} will not be altered under a swap operation. We can provide a counter-example where this does not hold: a particular choice of P-R box [54] (see also § 2.6.2). Consider a bipartite system, in which there are two normalised measurements $X_1 = \{e_1, \tilde{e}_1\}$ and $X_2 = \{e_2, \tilde{e}_2\}$, such that $e_1 = 1 - \tilde{e}_1$ and $e_2 = 1 - \tilde{e}_2$. A relatively concise representation of the state would be in terms of the statistics of the following effects:

$$s_{PR} = \begin{pmatrix} e_1 \otimes 1 \\ e_2 \otimes 1 \\ 1 \otimes e_1 \\ 1 \otimes e_2 \\ e_1 \otimes e_1 \\ e_1 \otimes e_2 \\ e_2 \otimes e_1 \\ e_2 \otimes e_2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \\ 0 \\ 1/2 \end{pmatrix}, \quad (7.14)$$

where we have taken a representation in which we explicitly write out the locally measurable statistics. (The probabilities associated with the other effects $\{e_i \otimes \tilde{e}_j\}$, $\{\tilde{e}_i \otimes e_j\}$ and $\{\tilde{e}_i \otimes \tilde{e}_j\}$ can all be determined from the above, assuming normalisation $(1 \otimes 1) \cdot s = 1$.) After some minor calculation, one verifies that $X_1 \otimes X_1$ and $X_2 \otimes X_2$ and $X_1 \otimes X_2$ are all correlated, but $X_2 \otimes X_1$ is anti-correlated. Despite the marginal states of A and B being manifestly identical, the swap operation (which exchanges $e_1 \otimes e_2 \leftrightarrow e_2 \otimes e_1$) will thus have an effect on s_{PR} . There is hence at least one measurement outcome $e_1 \otimes e_2$ whose statistics could reveal to an experimentalist that the systems had been exchanged.

We can also construct a separable quantum example

$$\rho = \frac{1}{3} \left(|0\rangle\langle 0| \otimes |+\rangle\langle +| + |+\rangle\langle +| \otimes |-\rangle\langle -| + |-\rangle\langle -| \otimes |0\rangle\langle 0| \right) \quad (7.15)$$

where $\text{Tr}_B(\rho) = \text{Tr}_A(\rho)$, and so no quantity measured locally on either side will reveal which is A and which is B ; but the correlations $\langle Z \otimes X \rangle = \frac{1}{3}$, whilst $\langle X \otimes Z \rangle = -\frac{1}{3}$ such that something observable changes if the subsystems are exchanged.

Following a similar pattern, even classical theory admits a counter-example for three-level systems (with possibilities 1, 2 and 3 associated with effects e_1 , e_2 and e_3) prepared in such a correlated way:

$$s = \frac{1}{3} \left(1 \otimes 2 + 2 \otimes 3 + 3 \otimes 1 \right), \quad (7.16)$$

in which the direction of permutation (1, 2, 3) can identify one side as A and the other as B since $(e_1 \otimes e_2) \cdot s \neq (e_2 \otimes e_1) \cdot s$.

7.2.2 Post-quantum controlled-swap

Let us now see if we can formulate a controlled-swap in the GPT framework. Suppose there are two systems A and B with identical state-spaces. Let the state $\mathbf{s}_{AB}$ describe the joint state of these two systems (encapsulating all statistics of measurements on A and B , as well as the correlation statistics between the two systems, and possibly also non-locally-tomographic parts). Let there also be a (different, in general) two-level system C whose state is written $\mathbf{s}_C$. Consider an initial state with no correlations between AB and C , such that we might write the tensor product:

$$\mathbf{s}_{ABC}^{\text{in}} = \mathbf{s}_{AB}^{\text{in}} \otimes \mathbf{s}_C^{\text{in}}. \quad (7.17)$$

Let the subsystems A and B be *indistinguishable* (according to definition 7.2) such that an unconditional swap on the first two subsystems has no effect on the state:

$$(T_{\text{swap}} \otimes \mathbb{1}) \mathbf{s}_{ABC}^{\text{in}} = (T_{\text{swap}} \mathbf{s}_{AB}) \otimes \mathbb{1} = \mathbf{s}_{ABC}^{\text{in}}. \quad (7.18)$$

When expressed in the GPT framework, even if the swap operation induces potentially different global phases (such as it would for symmetric and anti-symmetric quantum states $\mathbf{s}_{AB}$), this is unmeasurable, and the operational state $\mathbf{s}_{ABC}^{\text{in}}$ will not be changed by an unconditional swap.

Again, taking our lead from the quantum scenario, let us now we propose a new reversible transformation T_{CSwap} , in which conditionally swap A and B based on the state of C . We do not allow this transformation to change the “should I swap?” measurement of C , so that a series of swaps conditioned on the same subsystem C , either all swap, or none swap. This is natural restriction for the sort of systems we are considering (see fig. 7.1), in which there may be spatial separation between the regions where swapping occurs and where it does not, and so we do not wish the “swap” to cause the particles to jump from one region into the other.

We require that when AB is indistinguishable, the application of T_{CSwap} cannot be detected by any measurement or correlation in A and B , such that the reduced states of AB are the same before and after the controlled-swap: $\mathbf{s}_{AB}^{\text{in}} = \mathbf{s}_{AB}^{\text{out}}$. (The alternative would be that although unconditionally swapping and doing nothing have no effect on indistinguishable $\mathbf{s}_{AB}$, some possibly coherent combination of the two could induce $\mathbf{s}_{AB}$ to change.)

Let us particularly consider the case where $\mathbf{s}_{AB}$ is initially prepared in a *pure* indistinguishable state. This allows us to use theorem 2 in [129] (which holds regardless of whether the theory is locally tomographic), which states that if one reduced state of a bipartite system is pure, the entire system must be in a product state. From this, we infer that the output state must also be a product state, because $\mathbf{s}_{AB}$ is unchanged by

the operation and so remains pure. The action of the control-swap operation on such a state will be as follows:

$$\begin{aligned} T_{\text{CSwap}}(\mathbf{s}_{AB}^{\text{in}} \otimes \mathbf{s}_C^{\text{in}}) &= \mathbf{s}_{AB}^{\text{out}} \otimes \mathbf{s}_C^{\text{out}} \\ &= \mathbf{s}_{AB}^{\text{in}} \otimes \mathbf{s}_C^{\text{out}}, \end{aligned} \quad (7.19)$$

where the only observable effect (for this input state) is to transform $\mathbf{s}_C^{\text{in}} \rightarrow \mathbf{s}_C^{\text{out}}$. Moreover, if T_{CSwap} is a reversible operation, then $\mathbf{s}_C^{\text{out}}$ will be a pure state if $\mathbf{s}_C^{\text{in}}$ is also pure. As T_{CSwap} does not change the “should I swap?” measurement of C , this means the resultant state of T_{CSwap} when $\mathbf{s}_{AB}$ is indistinguishable is the same state that would be obtained if the transformation $\mathbb{1} \oplus T_\Phi$ were applied³, where T_Φ is an element of the phase group associated with the “should I swap?” measurement on C .

Thus if there is any “phase kickback” such as there is in quantum theory, the only place in this system where it could be observed will be on the control system C . This holds regardless of whether the joint system of AB and C is locally tomographic. We synthesise the above discussion in the following theorem:

Theorem 7.1. *The observable effect of a controlled exchange of indistinguishable systems manifests as a phase transformation associated with the control bit.*

The precise relationship between the types of correlations within the system AB , and whether exchanging labels always ‘kicks back’ a phase to the phase group of the control bit, has not yet been answered. Thus, for the remainder of this discussion, we will only make claims about what *could* be observed, should such a mechanism be possible.

7.3 Generalised particle types

In the same way we that different particle types induce different complex phases on the quantum control bit, we consider if there might be an associated particle type for every different element of the (generalised) phase group on the post-quantum control bit.

In quantum theory, by measuring the phase shift, we might identify particles as *bosonic*, *fermionic* or *anyonic*. We thus present generalisations of these definitions, in the context of arbitrary GPTs (noting that they will reduce to the quantum definition for the special case of quantum theory).

³For a non-indistinguishable state $\mathbf{s}_{AB}$, this statement will not be true in general.

Definition 7.3: Bosonic swap

If the observable effect of a controlled-swap on indistinguishable particles is to act on the control bit with $\mathbb{1}$, then the swap behaviour is identified as **bosonic**.

This is the case for quantum states where $T_{\text{swap}} : |\Psi(\xi, \chi)\rangle \rightarrow |\Psi(\chi, \xi)\rangle$.

Definition 7.4: Fermionic swap

If the observable effect of a controlled-swap on indistinguishable particles is to act on the control bit with a self inverse operation T_f such that $T_f T_f = \mathbb{1}$ but $T_f \neq \mathbb{1}$, then the swap behaviour is identified as **fermionic**.

This is the case for quantum states where $T_{\text{swap}} : |\Psi(\xi, \chi)\rangle \rightarrow -|\Psi(\chi, \xi)\rangle$.

Definition 7.5: Anyonic swap

If the swap behaviour is neither bosonic or fermionic, it is **anyonic**.

This is the case for quantum states where $T_{\text{swap}} : |\Psi(\xi, \chi)\rangle \rightarrow e^{i\alpha} |\Psi(\chi, \xi)\rangle$ for $\alpha \in [0, 2\pi)$ and $\alpha \neq 0, \pi$.

7.3.1 Example: classical theory

Suppose we attempt to turn the above definition on indistinguishable classical particles. In the information-theoretic approach of the GPT framework, definition 7.2 still make sense. As classical systems only have a single measurement, the control bit for a controlled-swap can only execute a perfect swap, no action, or a probabilistic mixture of the two. The only phase operation on the classical control bit is $\mathbb{1}$, and hence the only swap behaviour that could be identified would be *bosonic* according to the above definitions.

For this reason (at least in the setup we are considering), it is reasonable to consider bosonic behaviour as the simplest class of behaviour when indistinguishable particles are exchanged. In particular as $\mathbb{1}$ is an element of every phase group, then it is plausible that bosonic swap behaviour may be identified in every theory (although whether it is actually *admitted* will depend exactly on which swap operations can be found!)

7.3.2 Example: ball state-spaces

Ball state-spaces (as discussed in § 6.4) provide a natural generalisation of two-level quantum systems, and so we view n -balls as a likely contender for post-quantum particles.

The $n = 1$ case reduces to classical bits, which (as discussed above) only admits bosonic swap behaviour.

The $n = 3$ case is exactly complex quantum theory, with the phase parametrised by a single angle ϕ . There is one bosonic swap corresponding to $\phi = 0$, one fermionic swap given by $\phi = \pi$, and all other ϕ correspond to anyons.

A $n = 2$ ball with phase group $O(1) \cong Z_2$ admits two reversible controlled-swaps operations. The first is associated with $\mathbb{1}$ and is hence bosonic. The second is associated with the phase group element encapsulating reflection across the diameter of the circle, which is a self-inverse element and hence fermionic. This particular swap operation might be viewed as somewhat unphysical as it does not have time-continuity (as discussed in § 6.4.1): there is no way to apply a fractional portion of the fermionic controlled-swap operation. The phase group is a two-element finite group, and thus we have exhausted the possible swap behaviours: there are no anyonic swaps.

Consider the $n = 4$ case. The existence of the reversible coherent CSWAP operation is not trivially answered, as it depends very much on what states are in the tri-partite state-space formed by taking the tensor product of three 4-ball state-spaces, which is itself a mathematically complex question (see the discussion at the start of § 2.6). For the purpose of this discussion, however, suppose a transformation satisfying the requirements of § 7.2.2 does exist. Taking the phase group as $SO(3)$ (by disallowing transformations that are not time-continuous from the full automorphism group $O(4)$), we still have a single bosonic swap operation $\mathbb{1}$. However, there is now in fact a set of transformations $R \in SO(3)$ satisfying $RR = \mathbb{1}$, $R \neq \mathbb{1}$: a 180° rotation about *any* axis will be self inverse. Thus we will have a family of fermionic swaps.

Suppose we had swap operations associated with the three elements of $SO(3)$ associated with 180° rotations around the X , Y or Z axis of the sphere: there will exist 4-ball control states with statistics that change in different ways, depending on which of these apply: resulting in a non-equivalent classes of fermions. Moreover, these elements are non-abelian. This gives rise to what we might term *non-abelian fermions*. This non-abelianity is somewhat different from the braiding that can occur with quantum “non-abelian anyons”, as it has not arisen from any topological constraint in the physical space in which the exchange takes place (see e.g. [125, 126, 130, 131]), but rather as a consequence of the state-space of the control bit. (Consider for example if there were two X type fermions and two Y type fermions, and a control bit initially in a coherent state. The final state of this control bit would be different if it first triggered

the X swap then the Y swap, than if it first triggered the Y swap and then the X .)

On the other hand, if for the 4-ball we allowed only the transitive $U(2)$ subgroup of dynamics, we recover the $U(1)$ quantum phase group, and the observable swap types would be exactly as those in complex quantum theory.

Higher dimensional ball state-spaces will exhibit similar non-abelian behaviour for the transitive groups of dynamics, in which there will be in general more than one type of fermion, and the swaps will not commute.

Note, however, that the existence of multiple types of fermion does not in itself guarantee non-abelian behaviour. Consider the case of toroidal dynamics $\bigoplus^n SO(2)$, e.g. for $n = 5$ the phase group $SO(2) \oplus SO(2)$. The group of self-inverse elements here is $Z_2 \oplus Z_2$, in which $\mathbb{1} \oplus \mathbb{1}$ is associated with a bosonic swap, and $X \oplus \mathbb{1}$, $\mathbb{1} \oplus X$ and $X \oplus X$ give three observably different fermionic swap behaviours. These operations all commute with each other.

7.3.3 Example: ‘box-world’

Not every GPT admits a controlled-swap transformation that operates exactly according to the description in § 7.2.2. Consider ‘box-world’, which admits any (locally tomographic) set of non-signalling statistics, and so has gbits as its single-system state-spaces. Gross and co-authors [50] have demonstrated that box-world does not admit any reversible dynamics other than the simple unconditional permutations of measurements, measurement outcomes, or subsystems.

This rules out controlled-operations like the controlled-swap as a reversible dynamic. We can see that the controlled-swap is intrinsically correlating by considering an initial product state $s_A \otimes s_B \otimes s_C$, where $s_A \neq s_B$, and the “should I swap” measurement on s_C is not in state of definitely swapping, or definitely not swapping. The effect of the controlled-swap will then correlate the “should I swap?” measurement with $s_A \otimes s_B$ vs. $s_B \otimes s_A$.

In box-world, if the “should I swap?” measurement (we will label this Z) is one of the usual measurements along the edge of the cube for the gbit, then there are no pure states that exist as a superposition of both swapping and not swapping (for gbits, all pure states are the corners which will be located on the planes associated with definite values of every measurement including Z). There are no pure superposition states equivalent to (say) $|+\rangle$ in quantum theory, in which both outcomes of the Z measurement have support, but the state is pure. For gbits, the other complementary measurements may have definite outcomes, even when Z is also known. It seems therefore feasible that a phase kickback associated with a swapping operation (which is only visible for super-position states of the control bit in quantum physics) could manifest itself in a control gbits, even when the control bit is in a definite state of the

“should I swap?” measurement.

Even taking into account the stringent requirements shown in [50], an unconditional swap transformation is permissible as it corresponds to a permutation of subsystem labels. We write this as $T_{\text{swap}} : \mathbf{s}_{AB} \rightarrow \mathbf{s}_{BA}$. Likewise, not swapping (i.e. the identity transformation) is also always permitted. We thus consider if there is a *irreversible* controlled-swap T_{irr} which that exhibits suitable swapping behaviour. The simplest such operation would be a classical mixture of swapping and not swapping:

$$T_{\text{irr}} : \mathbf{s}_{AB} \otimes \mathbf{s}_C \rightarrow (\mathbf{z}^+ \cdot \mathbf{s}_C) \mathbf{s}_{BA} \otimes \mathbf{s}_C^+ + (\mathbf{z}^- \cdot \mathbf{s}_C) \mathbf{s}_{AB} \otimes \mathbf{s}_C^- \quad (7.20)$$

in which effect $\mathbf{z}^+$ is associated with swapping, and $\mathbf{z}^-$ with not swapping, and $\mathbf{s}_C^+$ is a state with all the same statistics as $\mathbf{s}_C$ except $\mathbf{z}^+ \cdot \mathbf{s}_C^+ = 1$ (and likewise $\mathbf{s}_C^-$ is as $\mathbf{s}_C$ but $\mathbf{s}_C^- \cdot \mathbf{s}_C^- = 1$). Because the two addends in the sum are weighted by the Z measurement outcomes of $\mathbf{s}_C$, the reduced state of C after the swap is $\mathbf{s}_C$.

As the two addends of eq. (7.20) are both allowed states in box-world, the convex combination of them should also be admissible. Furthermore, T_{irr} is a linear map. Thus, as a linear map from valid states to valid states, T_{irr} is a plausible transformation.

More interestingly, we can generalise T_{irr} , such that it has a non-trivial effect on the reduced state $\mathbf{s}_C$:

$$T_{\text{irr}} : \mathbf{s}_{AB} \otimes \mathbf{s}_C \rightarrow (\mathbf{z}^+ \cdot \mathbf{s}_C) \mathbf{s}_{BA} \otimes \mathbf{t}_C^+ + (\mathbf{z}^- \cdot \mathbf{s}_C) \mathbf{s}_{AB} \otimes \mathbf{s}_C^-, \quad (7.21)$$

where $\mathbf{t}_C = T_\Phi \mathbf{s}_C$ and (as above) $\mathbf{t}_C^+$ and shares exactly the same statistics as $\mathbf{t}$, save for the outcomes of the Z measurement. Note that the phase element is only applied to C in the instance where the particles are swapped. Take the case now when A and B are in an indistinguishable state. The observable effect of the transformation is now only on the control bit $\mathbf{s}_C \rightarrow (\mathbf{z}^+ \cdot \mathbf{s}_C) \mathbf{t}_C^+ + (\mathbf{z}^- \cdot \mathbf{s}_C) \mathbf{s}_C^-$.

When $T_\Phi = \mathbb{1}$, there are no observable changes, and the control-swap operation could be repeated an arbitrary number of times without mixing any $\mathbf{s}_C$. However, suppose T_Φ takes is a non-trivial element of the gbit phase group (see § 3.3.3), corresponding to the rotations of a hypercube of one dimension less than the number of measurements in the gbit. When $\mathbf{z}^+ \cdot \mathbf{s}_C = 1$, we can directly observe a different behaviour for each distinct element in the phase group of T_Φ . In this case, pure states of C are mapped to pure states of C , and so $\mathbf{s}_C$ will not be mixed, allowing for the effect of the transformation to be deterministically reversed for this particular class of input state. However, if $0 < \mathbf{z}^+ \cdot \mathbf{s}_C < 1$, then for $T_\Phi \neq \mathbb{1}$, system C will become mixed, making it operationally trickier to measure what happened to C .

This behaviour is in contrast with a quantum swap, where the phase kickback is most measurable for control bits in the equiprobable superposition $|\psi\rangle_C = \frac{1}{\sqrt{2}} (|Z^+\rangle + e^{i\phi}|Z^-\rangle)$. In the quantum case, the phase shift might also be directly associated with the $+1$

eigenstate of the control bit $|Z^+\rangle$, but the uncertainty principle limits how observable this phase is. Crucially, the lack of uncertainty principle on gbits makes the observation of phase changes on the control bit plausible, even for definite states of swapping.

Let us see what sort of particles this alternative construction could reveal. As usual, we will always find bosonic swap behaviour, associated with the element $\mathbb{1}$. The number of fermions depends on dimension of the control bit, and which measurement is taken as the control. For 2-in 2-out gbits, we may find a single fermion (associated with swapping $X \rightarrow -X$) (or an equivalent single flip for any other binary measurement selected). For 3-in 2-out gbits, taking the phase group along one of the cube axes, we have the automorphisms of the square. This gives us two fermions, 180° rotations, and flips. There are also four anyonic swaps, corresponding to rotations 90° either way, possibly followed by a flip. (If flips are not allowed, we only have the one fermionic swap, and two anyonic swaps.) For higher dimensions, we find a finite number of fermionic and anyonic swaps; and in general these fermionic swap operations do not commute.

7.3.4 Example: Spekkens' toy model

As a final example, we consider Spekkens' toy model. The phase group of a Spekkens' bit is $Z_2 \oplus Z_2$ (see §3.3.4), and so providing an appropriate operation exists, there would be the potential to observe four different types of particle behaviour (one bosonic, and three fermionic). Although these each of the phase operations will transform from one superposition state into one of the other three in the set $\left(\begin{smallmatrix} \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \end{smallmatrix} \text{ or } \begin{smallmatrix} \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \end{smallmatrix} \right)$, because these states are not orthogonal, repeated runs of the experiment would be required to distinguish which type of behaviour was being observed.

Whilst it is possible to construct an unconditional swap of a two-bit system (simply by permuting the labels of subsystems), the epistemic restriction unfortunately prevents the formulation of a reversible controlled-swap model for a three-bit system in Spekkens' toy model. The proof is a lengthy digression from the theme of this chapter, and so is provided instead in appendix B

Thus, without a controlled-swap gate, the experimental set-up detailed in this chapter will not be possible. We remark that the non-existence of a controlled-swap gate also means that Spekkens' toy model is not a suitable candidate to simulate universal quantum computing, even though it admits a controlled-not transformation and a broad set of single-bit dynamics.

7.4 Conclusion and outlook

In this chapter, we discussed the operational definition of indistinguishability in the GPT framework and outlined a hypothetical set-up that could allow us to discriminate between different types of particles in the GPT framework beyond the bosons and fermions of quantum theory.

The set-up proposed here works by inducing a phase on a control bit, and so the types of particle behaviour we could observe are a function of the control bit state-space. Thus, even if there are post-quantum particle types, if we can only observe their exchange behaviour in a device controlled by a quantum degree of freedom (qubit), then we would only ever be able to observe the quantum particle types. (This is not as unusual as it at first sounds: consider exchanging quantum particles using a classical control bit; as there is no coherence, there will be no observable difference in statistics between the two species of particle, and hence the quantum species will not be identifiable on a classical control bit.)

Moreover, the construction provided in this chapter discusses *what* could be observed in a particular scenario, but does not directly address the concept of different types of symmetrisation, which accounts for *why* some quantum states are bosonic and some are fermionic. This poses a natural open question about how to abstract the concept of symmetrisation in terms of measurement statistics (presumably correlations) between two general systems in the GPT framework. In particular, if a theory admits of more exotic possibilities⁴ than just symmetrisation and one type of anti-symmetrisation; can we construct a controlled-swap operation for that theory such that different symmetrisation classes result in different elements of the phase group being induced on the control bit?

If this can be done, one would hope to find a controlled-swap transformation such that for each symmetry class Φ , the effect of the transformation on all indistinguishable states in this class is as if $\mathbb{1}_{AB} \oplus T_\Phi$ were applied; much as how the usual quantum CSWAP behaves as $\mathbb{1}_{AB} \oplus \mathbb{1}$ for symmetric bosons, and as $\mathbb{1}_{AB} \oplus Z$ for anti-symmetric fermions.

⁴We suggest that each possible symmetry class could be identified as distinct by a rule that disconnects indistinguishable states in one class from those in another, e.g. no transformation of the form $\otimes^N T$ (where T is an allowed transformation on a single system) acting on a state of N indistinguishable systems should be able to map states between classes.

Chapter 8

Conclusion

In this thesis, we have shown how phase can be taken beyond quantum Hilbert spaces and into the framework of generalised probabilistic theories (chapter 3). We have seen that we can describe an interference experiment in almost any non-classical theory (chapter 4), but considerations of locality limit us to a much smaller class of physically-viable theories (chapters 5 and 6). The GPT formulation of phase also allowed us to describe the observable statistics for post-quantum particles (chapter 7).

Alongside the open questions posed in the conclusions of each chapter, one broad outlook for future research would be to use the concepts laid down in this thesis as the basis of a general derivation of quantum theory in terms of axioms solely related to locality. For example, no-signalling (intrinsic to the GPT framework), branch locality (chapter 5), and consistency with relativity of simultaneity (chapter 6) could all be posited as fundamental axioms associated naturally with the concept of “over here vs. over there”. If other axioms of a similar flavour could be found, it might prove for a satisfying physically-motivated derivation of quantum theory, as opposed to one that follows from abstract information-theoretical concepts. Building this sort of locality directly into quantum theory could pave the way towards a consistent theory of quantum gravity (see e.g. [132]).

Another direction of extension would be to consider how phase appears in an alternative set-up: that of energy (i.e. the *Hamiltonian*). In quantum theory, a closed system with energies $\{E_i\}$ associated with each level evolves in time t according to the unitary operation $U = \sum_i e^{iE_i t} |E_i\rangle\langle E_i|$. This has the form of a phase transformation with respect to the energy measurement; and so the concept might naturally generalise to arbitrary theories¹. The locality principles deployed in this thesis would no longer be directly applicable in this context, but one might be able to impose another restriction along the lines of “all degrees of freedom of the same dimension behave in the same operational manner” (and this in itself might be motivated by a bolder claim that ultimately the *only* thing that can be measured in an experiment reduces to a position of some kind [133]).

¹Cf. [17], which shows that only the toroidal phase group $\bigoplus \text{U}(1)$ of complex quantum theory has a one-to-one correspondence between outcomes $\{E_i\}$ and the degrees of freedom in the choice of transformation, specifically within the set of state-spaces described by a Jordan algebra.

Finally, we remark that although phase is not unique to quantum theory, the study of it in generalised probabilistic theories has allowed us to better understand its quantum manifestation. In particular we have clarified how the features of quantum theory collude to form a good description of reality: the quantum uncertainty principle censors the measurement of statistics associated with dynamics that would otherwise violate locality, and the abelian nature of quantum phase dynamics ensures that interferometry experiments are compatible with relativity of simultaneity. One hopes that continued research into this topic will yield similar insights into the fundamental nature of reality.

Appendix A

Uncertainty in quantum theory

The restriction on the values that complementary measurements in quantum theory can take arises as a direct consequence of a geometric property of Hilbert spaces known as *the Cauchy-Schwarz inequality*. This is an inequality for inner products such that

$$\langle \xi | \xi \rangle \langle \chi | \chi \rangle \geq |\langle \xi | \chi \rangle|^2 \quad (\text{A.1})$$

which we can use to derive a bound on the product of variances of measurement operators.

First note that for all pure quantum states $|\psi\rangle$, and two operators A and B , the Cauchy-Schwarz inequality implies that $|\langle \psi | AB | \psi \rangle|^2 \leq \langle \psi | A^\dagger A | \psi \rangle \langle \psi | B^\dagger B | \psi \rangle$, which when A and B are Hermitian further reduces to

$$|\langle \psi | AB | \psi \rangle|^2 \leq \langle \psi | A^2 | \psi \rangle \langle \psi | B^2 | \psi \rangle. \quad (\text{A.2})$$

For Hermitian A and B , $(\langle \psi | AB | \psi \rangle)^\dagger = \langle \psi | BA | \psi \rangle = \langle \psi | AB | \psi \rangle^*$, and so $\langle \psi | (AB + BA) | \psi \rangle$ is a real number equal to twice the real part of $\langle \psi | AB | \psi \rangle$. Similarly, $\langle \psi | (AB - BA) | \psi \rangle$ is purely imaginary, equal to twice the imaginary part of $\langle \psi | AB | \psi \rangle$. Using $|\bullet|^2 = \Re(\bullet)^2 + \Im(\bullet)^2$, we thus find that $4|\langle \psi | AB | \psi \rangle|^2 = |\langle \psi | (AB + BA) | \psi \rangle|^2 + |\langle \psi | (AB - BA) | \psi \rangle|^2$. Substituting this expression into the Cauchy-Schwarz inequality have

$$|\langle \psi | [A, B] | \psi \rangle|^2 + |\langle \psi | \{A, B\} | \psi \rangle|^2 \leq 4\langle \psi | A^2 | \psi \rangle \langle \psi | B^2 | \psi \rangle, \quad (\text{A.3})$$

where we have written the *commutator* $[A, B] = AB - BA$ and the *anti-commutator* $\{A, B\} = AB + BA$. If we make the substitutions $A = X - \langle X \rangle$ and $B = Y - \langle Y \rangle$, it may be verified that $[A, B] = [X, Y]$, and $\{A, B\} = \{X, Y\} - 2(X\langle Y \rangle + Y\langle X \rangle) + 2\langle X \rangle \langle Y \rangle$, such that $\langle \{A, B\} \rangle = \langle \{X, Y\} \rangle - 2\langle X \rangle \langle Y \rangle$. Because $|\langle \psi | \{A, B\} | \psi \rangle|^2$ is never negative, from eq. (A.3) we can write

$$\begin{aligned} |\langle [X, Y] \rangle|^2 &\leq 4(\langle X^2 \rangle - \langle X \rangle^2)(\langle Y^2 \rangle - \langle Y \rangle^2) \\ \frac{1}{4}|\langle [X, Y] \rangle|^2 &\leq \Delta X^2 \Delta Y^2 \end{aligned} \quad (\text{A.4})$$

where we have written the *variance* of X as $\Delta X^2 = (\langle X^2 \rangle - \langle X \rangle^2)$ (and likewise for Y), and all expectation values $\langle \bullet \rangle$ are with respect to the same pure state $|\psi\rangle$. This is the

usual expression of quantum uncertainty arising as a direct result of non-commutating operators.

In order to get a tighter inequality (which we will need to motivate the precise restriction on quantum state-spaces), we consider the substitution into eq. (A.3) again, this time including the anti-commutation term to precisely write

$$|\langle[X, Y]\rangle|^2 + |\langle\{X, Y\}\rangle - 2\langle X\rangle\langle Y\rangle|^2 \leq 4(\langle X^2\rangle - \langle X\rangle^2)(\langle Y^2\rangle - \langle Y\rangle^2), \quad (\text{A.5})$$

where this expression is *exactly* the Cauchy-Schwarz bound, as it can be applied to the expectation values of any two Hermitian operators X and Y .

Let us now see how this inequality can be used to demonstrate that a qubit has a spherical state-space when defined in terms of the Pauli operators

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (\text{A.6})$$

We substitute any two of the Pauli matrices (say σ_x and σ_y) into eq. (A.5) for X and Y . As $\sigma_x^2 = \sigma_y^2 = \mathbb{1}$, the variances on the right-hand side are given by $1 - \langle\sigma_x\rangle^2$ and $1 - \langle\sigma_y\rangle^2$, and noting from direct calculation that $[\sigma_x, \sigma_y] = 2i\sigma_z$ and $\{\sigma_x, \sigma_y\} = 0$ we may simplify the left-hand side to find

$$\langle\sigma_z\rangle^2 + \langle\sigma_x\rangle^2\langle\sigma_y\rangle^2 \leq (1 - \langle\sigma_x\rangle^2)(1 - \langle\sigma_y\rangle^2), \quad (\text{A.7})$$

which we re-arrange into

$$\langle\sigma_x\rangle^2 + \langle\sigma_y\rangle^2 + \langle\sigma_z\rangle^2 \leq 1. \quad (\text{A.8})$$

Thus, we see that the expectation values admissible for pure states are bounded by a sphere of unit radius. As pure states form the points at the edge of the convex set, all other admissible states must be within this sphere. Hence the state-space of a qubit is bounded by a three-dimensional sphere.

Appendix B

Non-existence of controlled-swap in Spekkens' toy model

First we note with the following lemma that there is a unique two-bit swap operation in Spekkens' toy model.

Lemma B.1. *The only valid swap operation in two bit Spekkens' toy model, that satisfies definition 7.1 for every valid epistemic state, corresponds to the unique permutation on the ontic states where $\mathbf{o}_A \otimes \mathbf{o}_B \rightarrow \mathbf{o}_B \otimes \mathbf{o}_A$ for all $\mathbf{o}_A, \mathbf{o}_B \in \{\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}, \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix}, \begin{smallmatrix} \blacksquare & \square \\ \blacksquare & \blacksquare \end{smallmatrix}, \begin{smallmatrix} \blacksquare & \square \\ \square & \blacksquare \end{smallmatrix}\}$.*

Proof. Visualising a two-bit system as a 4×4 grid, we see that this operation is equivalent to reflecting the grid across the diagonal. This is equivalent to relabelling subsystems A and B . It is obvious that exchanging ontic states of the subsystems is sufficient to guarantee that epistemic states are also exchanged.

Proving that this operation is *unique* is slightly more involved. A swap transformation will take the epistemic state $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \rightarrow \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$. A permutation of ontic states that achieves this must map the subset of the four supported ontic states to itself. However, the same is true for the state $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix}$ which is also mapped to itself. As $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ is the only ontic state with support in both these cases, the only way for these two requirements to simultaneously hold true is if the ontic state $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$ is mapped to itself. Similar logic can be used to argue that $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix}$, $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \square \\ \blacksquare & \blacksquare \end{smallmatrix}$ and $\begin{smallmatrix} \blacksquare & \square \\ \blacksquare & \blacksquare \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \square \\ \square & \blacksquare \end{smallmatrix}$ are also mapped to themselves.

Once we have deduced that the diagonal ontic states do not change, considering again $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$, we see that the off-diagonal ontic state $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix}$ must then either map to itself, or to the state in which the subsystem labels are exchanged, $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \square \\ \blacksquare & \blacksquare \end{smallmatrix}$. The same argument can be made for every off-diagonal ontic state in the grid.

Consider now when the two subsystems are now in a product state of two orthogonal states, such as $\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \square \\ \square & \blacksquare \end{smallmatrix}$ which will be mapped to $\begin{smallmatrix} \blacksquare & \square \\ \square & \blacksquare \end{smallmatrix} \otimes \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \square \end{smallmatrix}$. The states before and after do not share any of the same ontic possibilities, and so the permutation induced by the swap can not leave any ontic state supported in this state unchanged. Every off-diagonal ontic state can be spanned by such a construction, and hence we see that no off diagonal ontic states may be mapped to themselves by a swap.

Thus, the only remaining possibility is that every off-diagonal ontic state is taken to the state on the opposite side. This defines a unique permutation of ontic states that corresponds to the swap operation. $\square$

Due to the strict epistemic restrictions on allowed states, a reversible controlled swap operation is not permitted, as we shall now prove:

Theorem B.2. *Spekkens' toy model does not admit a three-bit controlled swap.*

Proof. First, note that the combination of a valid reversible transformation with a valid reversible transformation is also a valid reversible transformation. Moreover, the combination of a valid reversible transformation with an invalid reversible transformation is still invalid. Hence, in a search for an acceptable controlled swap operation, there is no preferred “should I swap?” measurement, and so without loss of generality, consider the cases where if the control bit is in state $\blacksquare\blacksquare\blacksquare$, a swap occurs, and if it is in $\blacksquare\blacksquare\blacksquare$ it does not. Likewise, the unconditional permutation of subsystems is always allowed, and so we will choose subsystem C to be the control bit.

Any ontic state where C is $\blacksquare\blacksquare\blacksquare$ or $\blacksquare\blacksquare\blacksquare$ (the “upper two layers” in a $4 \times 4 \times 4$ cubic grid visualisation) must be taken to the state where the ontic values of A and B are exchanged, but there remains a freedom in choice of $C = \blacksquare\blacksquare\blacksquare$ or $C = \blacksquare\blacksquare\blacksquare$ in the output. Likewise any state where C is $\blacksquare\blacksquare\blacksquare$ or $\blacksquare\blacksquare\blacksquare$ (the “lower two layers” in our cubic visualisation) must not change its values for A and B , but C could be cycled between $\blacksquare\blacksquare\blacksquare$ or $\blacksquare\blacksquare\blacksquare$.

Thus, consider when AB is a product state of identical A and B , and C is in a superposition such as $\blacksquare\blacksquare\blacksquare$. If either all, or none, of the states in $C = \blacksquare\blacksquare\blacksquare$ are taken to $C = \blacksquare\blacksquare\blacksquare$, the output state projects into a valid state of C . Likewise, if either all or none of the states in $C = \blacksquare\blacksquare\blacksquare$ are taken to $C = \blacksquare\blacksquare\blacksquare$, the output state also projects into a valid state of C . Suppose however if only some of the states are mapped across. In order that the projection onto C be valid, we require that two of the four states in $\blacksquare\blacksquare\blacksquare$ are taken to $\blacksquare\blacksquare\blacksquare$, and two of the four in $\blacksquare\blacksquare\blacksquare$ are taken to $\blacksquare\blacksquare\blacksquare$.

By considering identical product states, we can place a further restriction on the swap operation, which is illustrated in fig. B.1. In the upper two layers, if there are two ontic possibilities that can ‘switch’ between layers, then they must either both share the same reduced ontic state for A and B , or both having differing A and B . If this is not the case, after projecting onto AC or BC , there will be some measurement on C where one outcome completely specifies A or B ; as the projection is a two-to-one map of ontic states. This is a problem independently of what happens in the bottom layer (even if the bottom also has a two-to-one map): then there will be one measurement of C that is guaranteed to totally determine which ontic state A (or B), because for this particular input state, the transformation can not map any ontic state of A to a state which did not have support in the initial epistemic state of A . A

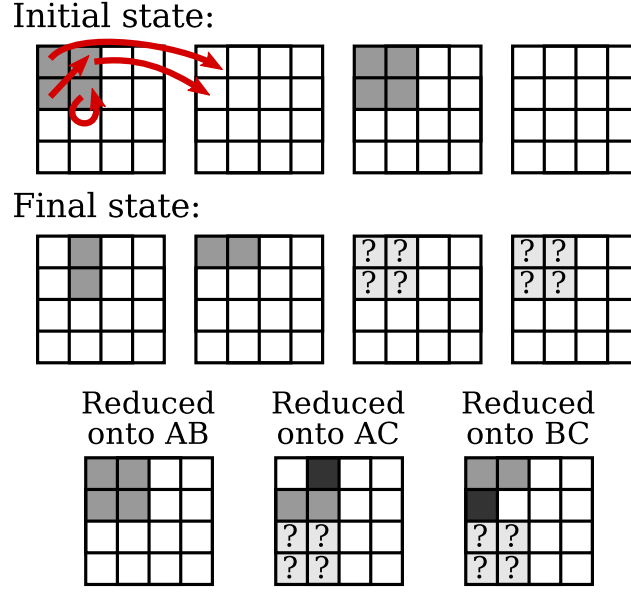


Figure B.1: Disallowed controlled swap. In this diagram (those following), the four grids associated with each three-bit system represent the ontic possibilities of AB for each of $C = \blacksquare\blacksquare\blacksquare$, $C = \blacksquare\blacksquare\square$, $C = \blacksquare\square\blacksquare$, and $C = \square\square\blacksquare$ from left to right, respectively. Here, the product state $\blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare$ is mapped by a candidate CSWAP gate. The action of the permutation on the four states in the top layer is shown. Regardless of which mapping is applied to the bottom four (such that they stay within the same reduced state of AB), the top two layers will be in an invalid state after the operation: this is because when projected onto AC or BC , as one value of C corresponds to two ontic states, whereas the other corresponds to just one.

similar argument can be made for the mapping of the bottom two layers. The effect of this restriction, when taken over all possible product states, is that the whole CSWAP transformation (not just its projection) must be symmetric under the exchange of A and B labels for all ontic states. Moreover, it can be deduced from the above (by considering different product states) that if any of the four diagonal elements $\{\blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare, \blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare, \blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare, \blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare\}$ are switched, then they must all be switched.

Now let us consider the operation where there is no switching in the upper two layers whatsoever. We can rule the validity of this out by a counter-example. The state $\blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare$ is mapped to $\blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare + ? \otimes \blacksquare\blacksquare\blacksquare + ? \otimes \blacksquare\blacksquare\blacksquare$, where the states ‘?’ in the bottom two layers are not fully determined, but are in the support of of the swapped configuration of AB , $\blacksquare\blacksquare\blacksquare \otimes \blacksquare\blacksquare\blacksquare$. Here, a single measurement of $\blacksquare\blacksquare\blacksquare$ vs. $\blacksquare\blacksquare\blacksquare$ on A or B could reveal precisely the ontic state of C . A similar argument can be made for transformations where *everything* switches. Independently, the same argument can be made on the lower two layers; and so both the top and the bottom two layers must have some states that switch, and some that do not. By considering the maximally mixed state of AB in a product with $C = \blacksquare\blacksquare\blacksquare$,

we see that in order to be mapped onto a state that projects correctly onto C , there must be exactly 8 switches in the upper two layers, and 8 switches in the lower two layers.

To summarise the restrictions on the CSWAP candidates so far: In the top two layers, there must be 8 choices of ontic state AB that result in the ontic state of C being switched. These choices are arranged in a pattern that is symmetric under the exchange of A and B . Moreover, if one diagonal element in a layer switches, all diagonal elements switch.

We can now exhaustively go through four classes of possible controlled-swap gates which satisfy the above, and eliminate them all by counter-example. As the choice to switch or not is a binary one taken over the grid of ontic states, we will find it convenient to also write the choice of which states are switched using the grid notation. We strongly advise the reader to view the accompanying diagrams provided for each case.

i) First (as illustrated in fig. B.2) consider the case where the 8 switches occur on the support of a product of two identical epistemic states as well as the product of two orthogonal identical epistemic states (e.g. states in the support of $\blacksquare\blacksquare \otimes \blacksquare\blacksquare + \blacksquare\blacksquare \otimes \blacksquare\blacksquare$). This will result in an invalid mapping on states in a product of one of these states with a non-orthogonal alternative state.

ii) Let us now consider the opposite scenario (as illustrated in fig. B.3), where the only ontic states that are switched in the top layer are in the support of a product of two orthogonal states (e.g. states in the support of $\blacksquare\blacksquare \otimes \blacksquare\blacksquare + \blacksquare\blacksquare \otimes \blacksquare\blacksquare$). To find a counter example for this, input a state which is a product of two states which are not orthogonal to each other, nor orthogonal (or identical to) the states which would all be switched to the other layer.

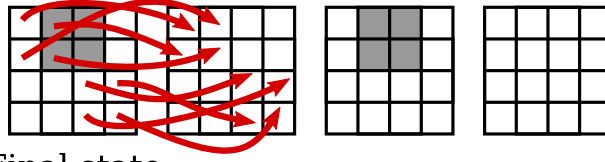
iii) Suppose now that the switches occur when just one product of identical states has complete support. This necessitates that all diagonal states switch, and so taken along with the symmetry, there is only one free choice (of three possibilities) of which element to switch remaining. Whatever choice we make for this, we can create another product state such that only one ontic possibility in the top two layers is switched. This gives us an invalid output and rules out such a swap (as illustrated fig. B.5).

iv) Finally, if no diagonal elements are switched, but the pattern of switching does not correspond to a product state of orthogonal elements, then it will always be possible (shown by exhaustion in fig. B.5) to construct some product state of A and B such that 3 ontic possibilities are mapped to $C = \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$, and only 1 remains in $C = \begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}$. This output will be invalid, as it can not correctly project onto C .

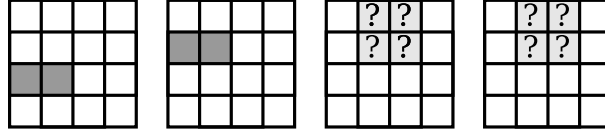
Thus, we have exhausted every possible mapping and failed to find a controlled swap operation in Spekkens' toy model.

□

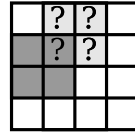
Initial state:



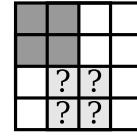
Final state:



Reduced
onto AB



Reduced
onto AC



Reduced
onto BC

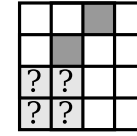
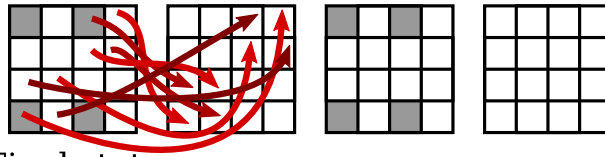
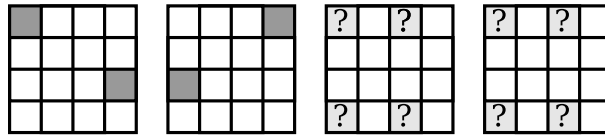


Figure B.2: Disallowed controlled swap for case i. The state $\begin{bmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{bmatrix} \otimes \begin{bmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{bmatrix}$ is mapped by a candidate CSWAP gate. The states in the top two layers which are switched are shown by red arrows. Regardless of which mapping is applied to the bottom four (such that they stay within the same reduced state of AB), the projection onto one of AC or BC results in a reduced state where one measurement of A (or B) could potentially reveal the exact ontic state of C.

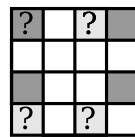
Initial state:



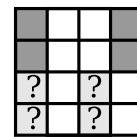
Final state:



Reduced
onto AB



Reduced
onto AC



Reduced
onto BC

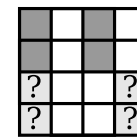


Figure B.3: Disallowed controlled swap for case ii. The state $\begin{bmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{bmatrix} \otimes \begin{bmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{bmatrix}$ is mapped by a candidate CSWAP gate. The states in the top two layers which are switched are shown by red arrows. Regardless of which mapping is applied to the bottom four (such that they stay within the same reduced state of AB), the projection onto at least one of AC or BC is reduced state which can not itself be correctly projected onto C.

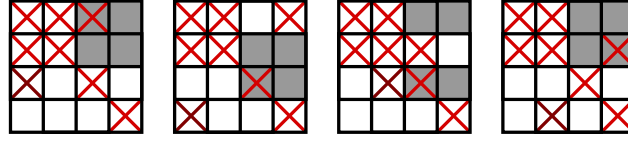


Figure B.4: Disallowed swap behaviour for case iii. There are 4 possible symmetrical choices of 8 states to swap after fixing one identical product (here $\blacksquare \otimes \blacksquare$), and including the remaining diagonal elements. Elements that are mapped to another layer are marked with a red cross. For each choice, a product state can be found (drawn in grey) such that one ontic possibilities is switched, and three are not. This results in an invalid projection onto C .

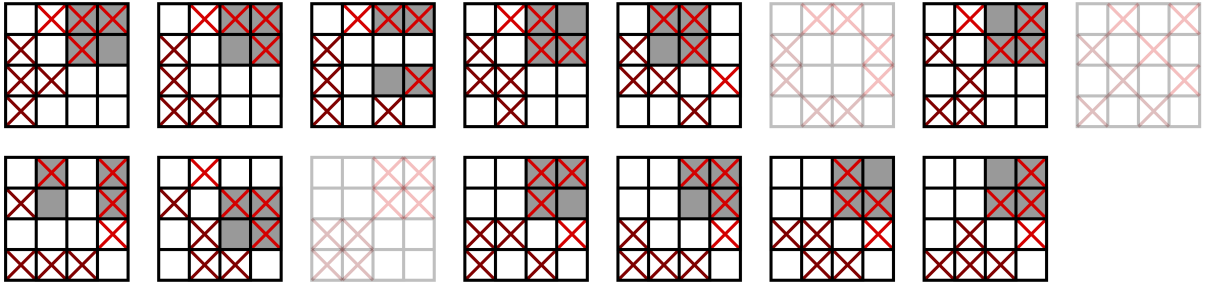


Figure B.5: Disallowed swap behaviour for case iv. There are 4C_6 possible symmetrical choices of 8 states to swap which contain no diagonal elements. Elements that are mapped to another layer are marked with a red cross. Three of these correspond to case ii, and are drawn (in faint) for completeness. For the rest of them, a product state can be found (drawn in grey) such that three ontic possibilities are switched, and one is not.

Bibliography

- [1] **A. J. P. Garner, O. C. O. Dahlsten, Y. Nakata, M. Murao, and V. Vedral.** *A framework for phase and interference in generalized probabilistic theories.* New Journal of Physics, 15(9):093044, September 2013.
- [2] **O. C. O. Dahlsten, A. J. P. Garner, and V. Vedral.** *The uncertainty principle enables non-classical dynamics in an interferometer.* Nature communications, 5 (4592), August 2014.
- [3] **A. J. P. Garner, M. P. Müller, and O. C. O. Dahlsten.** *The quantum bit from relativity of simultaneity on an interferometer.* arXiv:1412.7112, December 2014.
- [4] **O. C. O. Dahlsten, A. J. P. Garner, J. Thompson, M. Gu, and V. Vedral.** *Particle exchange in post-quantum theories.* arXiv:1307.2529, July 2013.
- [5] **N. D. Mermin.** *Could Feynman have said this?* Physics Today, 57(5):10–11, May 2004.
- [6] **W. K. Heisenberg.** *The physical principles of the quantum theory.* Dover, New York, NY, 1949 edition, 1930.
- [7] **E. W. Weisstein.** *Kelvin, Lord William Thomson (1824-1907), Eric Weisstein's World of Scientific Biography,* 1996. <http://scienceworld.wolfram.com/biography/Kelvin.html>.
- [8] **M. A. Nielsen and I. L. Chuang.** *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge, October 2000.
- [9] **L. Hardy.** *Quantum Theory From Five Reasonable Axioms.* arXiv:quant-ph/0101012, January 2001.
- [10] **L. Hardy.** *Probability theories in general and quantum theory in particular.* Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 34(3):381–393, September 2003.
- [11] **C. A. Fuchs.** *Quantum Mechanics as Quantum Information (and only a little more).* arXiv:quant-ph/0205039, May 2002.
- [12] **A. Zeilinger.** *A foundational principle for quantum mechanics.* Foundations of Physics, 29(4), April 1999.
- [13] **D. I. Fivel.** *Derivation of the Rules of Quantum Mechanics from Information-Theoretic Axioms.* arXiv:1010.5300, 2010.

- [14] **Ll. Masanes and M. P. Müller.** *A derivation of quantum theory from physical requirements.* New Journal of Physics, 13(6):63001, 2011.
- [15] **G. Chiribella, G. M. D’Ariano, and P. Perinotti.** *Informational axioms for quantum theory.* Foundations of Probability and Physics-6 AIP Conf. Proc., 270: 1424, 2012.
- [16] **G. Chiribella, G. M. D’Ariano, and P. Perinotti.** *Quantum Theory, Namely the Pure and Reversible Theory of Information.* Entropy, 14(12):1877–1893, October 2012.
- [17] **H. Barnum, M. P. Müller, and C. Ududec.** *Higher-order interference and single-system postulates characterizing quantum theory.* New Journal of Physics, 16(12):123029, December 2014.
- [18] **J. Barrett.** *Information processing in generalized probabilistic theories.* Phys. Rev. A, 75(3):32304, March 2007.
- [19] **S. Gasirowicz.** *Quantum Physics.* John Wiley & Sons, Hoboken, New Jersey, third edition, 2003.
- [20] **E. Schrödinger.** *Die gegenwärtige Situation in der Quantenmechanik.* Die Naturwissenschaften, 23(48):807–812, November 1935.
- [21] **D. Deutsch.** *Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 400(1818):97–117, July 1985.
- [22] **D. Deutsch and R. Jozsa.** *Rapid Solution of Problems by Quantum Computation.* Proc. R. Soc. Lond. A., 439(1907):553–558, 1992.
- [23] **D. R. Simon.** *On the power of quantum computation.* In Proceedings of the 35th Annual Symposium on Foundations of Computer Science, pages 116–123, November 1994.
- [24] **L. K. Grover.** *A fast quantum mechanical algorithm for database search.* In Proceedings of the twenty-eighth annual ACM symposium on theory of computing, STOC ’96, pages 212–219, New York, 1996. ACM.
- [25] **P. W. Shor.** *Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer.* SIAM J. Comput., 26(5):1484–1509, October 1997.
- [26] **R. Cleve, A. K. Ekert, C. Macchiavello, and M. Mosca.** *Quantum algorithms revisited.* Proc. R. Soc. Lond. A., 454(1969):339–354, January 1998.

- [27] **M. J. Bremner, R. Jozsa, and D. J. Shepherd.** *Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 467(2126):459–472, August 2010.
- [28] **R. W. Spekkens.** *Evidence for the epistemic view of quantum states: A toy theory.* Phys. Rev. A, 75(3):32110, March 2007.
- [29] **R. W. Spekkens.** *Quasi-quantization: classical statistical theories with an epistemic restriction.* arXiv:1409.5041, September 2014.
- [30] **A. Einstein.** *Relativity: The Special and General Theory.* Routledge, 1920.
- [31] **A. Hájek.** *Interpretations of Probability, Stanford Encyclopedia of Philosophy,* 2002. <http://plato.stanford.edu/archives/win2012/entries/probability-interpret/>.
- [32] **E. Seneta.** *Non-negative matrices and Markov chains.* Springer, New York, 2006.
- [33] **R. A. Horn and C. R. Johnson.** *Matrix Analysis.* Cambridge University Press, Cambridge, 1985.
- [34] **P. A. M. Dirac.** *A new notation for quantum mechanics.* Mathematical Proceedings of the Cambridge Philosophical Society, 35(03):416–418, July 1939.
- [35] **P. A. M. Dirac.** *The Principles of Quantum Mechanics.* Oxford University Press, Oxford, fourth edition, 1930.
- [36] **P. Szekeres.** *A Course in Modern Mathematical Physics: Groups, Hilbert Space and Differential Geometry.* Cambridge University Press, Cambridge, 2004.
- [37] **J. A. Wheeler and W. H. Zurek.** *Quantum Theory and Measurement.* Princeton University Press, Princeton, New Jersey, 1983.
- [38] **T. Heinosaari and M. Ziman.** *Quantum Theory: From Uncertainty to Entanglement.* Cambridge University Press, Cambridge, 2012.
- [39] **V. V. Gorbatsevich, E. B. Vinberg, and A. L. Onishchik.** *Foundations of Lie Theory and Lie Transformation Groups.* Springer-Verlag, Berlin, 1997.
- [40] **P. Ramond.** *Group Theory - A Physicist's Survey.* Cambridge University Press, Cambridge, 2010.
- [41] **M.-D. Choi.** *Completely positive linear maps on complex matrices.* Linear Algebra and its Applications, 10(3):285–290, June 1975.
- [42] **K. Kraus.** *States, Effects and Operations.* Springer-Verlag, Berlin, 1983.

- [43] **B. Dakić and Č. Brukner.** *Quantum Theory and Beyond: Is Entanglement Special?* Deep Beauty(...) Ed. H. Halvorson, Cambridge University Press, pages 365–392, 2011.
- [44] **H. Barnum and A. Wilce.** *Information Processing in Convex Operational Theories.* Electronic Notes in Theoretical Computer Science, 270(1):3–15, February 2011.
- [45] **C. Pfister.** *One simple postulate implies that every polytopic state space is classical.* Master’s thesis, ETH Zürich, 2012.
- [46] **P. Janotta and H. Hinrichsen.** *Generalized probability theories: what determines the structure of quantum theory?* Journal of Physics A: Mathematical and Theoretical, 47(32):323001, August 2014.
- [47] **A. Peres.** *Quantum Theory: Concepts and Methods.* Kluwer Academic Publishers, New York, 1995.
- [48] **P. G. L. Mana.** *Why can states and measurement outcomes be represented as vectors?* arXiv:quant-ph/0305117v3, May 2003.
- [49] **I. Bengtsson and K. Życzkowski.** *Geometry of Quantum States.* Cambridge University Press, Cambridge, 2006.
- [50] **D. Gross, M. Müller, R. Colbeck, and O. C. O. Dahlsten.** *All Reversible Dynamics in Maximally Nonlocal Theories are Trivial.* Phys. Rev. Lett., 104(8): 80402, 2010.
- [51] **G. Pisier.** *Introduction to Operator Space Theory.* Cambridge University Press, Cambridge, 2003.
- [52] **A. Einstein, B. Podolsky, and N. Rosen.** *Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?* Physical Review, 47(10): 777–780, May 1935.
- [53] **J. S. Bell.** *On the Problem of Hidden Variables in Quantum Mechanics.* Reviews of Modern Physics, 38(3):447–452, July 1966.
- [54] **S. Popescu and D. Rohrlich.** *Quantum nonlocality as an axiom.* Found. Phys., 24(3):379–385, 1994.
- [55] **V. Scarani.** *Feats, Features and Failures of the PR-box.* In AIP Conference Proceedings, volume 844, pages 309–320. AIP, March 2006.
- [56] **J. S. Bell.** *On the Eistein Podolsky Rosen paradox.* Physics, 1(3):195–200, 1964.

- [57] **J. Clauser, M. Horne, A. Shimony, and R. Holt.** *Proposed Experiment to Test Local Hidden-Variable Theories.* Physical Review Letters, 23(15):880–884, October 1969.
- [58] **B. S. Cirel'son.** *Quantum generalizations of Bell's inequality.* Letters in Mathematical Physics, 4(2):93–100, March 1980.
- [59] **J. S. Bell.** *Speakable And Unspeakable In Quantum Mechanics.* Cambridge University Press, Cambridge, 1987.
- [60] **L. A. Khalfin and B. S. Tsirelson.** *Quantum and Quasi-classical Analogs Of Bell Inequalities.* In **P. Lahti and P. Mittelstaedt** (editors), Symposium on the foundations of modern physics, pages 441–460. World Scientific Publishing Co., 1985.
- [61] **B. S. Tsirelson.** *Some results and problems on quantum Bell-type inequalities.* Hadronic Journal Supplement, 8:329–345, 1993.
- [62] **Li. Masanes, A. Acin, and N. Gisin.** *General properties of nonsignaling theories.* Physical Review A, 73(1):012112, January 2006.
- [63] **H. Barnum, J. Barrett, M. Leifer, and A. Wilce.** *Cloning and Broadcasting in Generic Probabilistic Theories.* arXiv:quant-ph/0611295, November 2006.
- [64] **H. Barnum, J. Barrett, M. Leifer, and A. Wilce.** *Generalized No-Broadcasting Theorem.* Phys. Rev. Lett., 99(24):240501, December 2007.
- [65] **S. Popescu.** *Nonlocality beyond quantum mechanics.* Nature Physics, 10(4):264–270, April 2014.
- [66] **W. van Dam.** *Nonlocality & Communication Complexity.* DPhil thesis, University of Oxford, Michaelmas Term 1999.
- [67] **W. van Dam.** *Implausible Consequences of Superstrong Nonlocality.* Natural Computing, 12(1):4, January 2013.
- [68] **G. Brassard, H. Buhrman, N. Linden, A. Méthot, A. Tapp, and F. Unger.** *Limit on Nonlocality in Any World in Which Communication Complexity Is Not Trivial.* Physical Review Letters, 96(25):250401, June 2006.
- [69] **R. Cleve, W. van Dam, M. Nielsen, and A. Tapp.** *Quantum entanglement and the communication complexity of the inner product function.* Theoretical Computer Science, 486:11–19, May 2013.
- [70] **H. Barnum, J. Barrett, M. Leifer, and A. Wilce.** *Teleportation in General Probabilistic Theories.* arXiv:0805.3553, May 2008.

- [71] **A. J. Short and J. Barrett.** *Strong nonlocality: a trade-off between states and measurements.* New Journal of Physics, 12(3):033034, March 2010.
- [72] **M. F. Pusey, J. Barrett, and T. Rudolph.** *On the reality of the quantum state.* Nature Physics, May 2012.
- [73] **M. F. Pusey.** *Stabilizer Notation for Spekkens' Toy Theory.* Foundations of Physics, 42(5):688–708, March 2012.
- [74] **S. J. van Enk.** *A Toy Model for Quantum Mechanics.* Foundations of Physics, 37(10):1447–1460, August 2007.
- [75] **P. Janotta and R. Lal.** *Generalized probabilistic theories without the no-restriction hypothesis.* Physical Review A, 87(5):052131, May 2013.
- [76] **A. Yu. Kitaev.** *Quantum measurements and the Abelian Stabilizer Problem.* arXiv:quant-ph/9511026, November 1995.
- [77] **D. Shepherd and M. J. Bremner.** *Temporally unstructured quantum computation.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 465(2105):1413–1439, February 2009.
- [78] **N. Linden, S. Popescu, A. J. Short, and A. Winter.** *Quantum mechanical evolution towards thermal equilibrium.* Physical Review E, 79(6):061103, June 2009.
- [79] **M. A. Schlosshauer.** *Decoherence - and the Quantum-To-Classical Transition.* Springer, 2007.
- [80] **K. F. Riley, M. P. Hobson, and S. J. Bence.** *Mathematical methods for physics and engineering.* Cambridge University Press, Cambridge, 1997.
- [81] **M. Kerin and D. Wraith.** *Homogeneous Metrics on Spheres.* Irish Math. Soc. Bulletin, 51:59–71, 2003.
- [82] **B. Coecke, B. Edwards, and R. W. Spekkens.** *Phase Groups and the Origin of Non-locality for Qubits.* Electronic Notes in Theoretical Computer Science, 270(2):15–36, 2011.
- [83] **I. D. Ivanovic.** *How to differentiate between non-orthogonal states.* Physics Letters A, 123(6):257–259, August 1987.
- [84] **A. Peres.** *How to differentiate between non-orthogonal states.* Physics Letters A, 128(1-2):19, March 1988.
- [85] **J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves.** *Symmetric informationally complete quantum measurements.* Journal of Mathematical Physics, 45(6):2171, May 2004.

- [86] **W. K. Wootters.** *Quantum Measurements and Finite Geometry*. Foundations of Physics, 36(1):112–126, February 2006.
- [87] **H. Barnum, J. Barrett, L. Orloff Clark, M. Leifer, R. W. Spekkens, N. Stepanik, A. Wilce, and R. Wilke.** *Entropy and information causality in general probabilistic theories*. New Journal of Physics, 12(3):033024, March 2010.
- [88] **M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski.** *Information causality as a physical principle*. Nature, 461:1101–1104, October 2009.
- [89] **C. Ududec, H. Barnum, and J. Emerson.** *Three Slit Experiments and the Structure of Quantum Theory*. Foundations of Physics, 41(3):396–405, February 2010.
- [90] **C. Ududec.** *Perspectives on the Formalism of Quantum Theory*. PhD thesis, University of Waterloo, 2012.
- [91] **R Sorkin.** *Quantum mechanics as quantum measure theory*. Mod. Phys. Lett. A, 9:3119, 1994.
- [92] **M. Barnett, F. Dowker, and D. Rideout.** *Popescu-Rohrlich Boxes in Quantum Measure Theory*. J. Phys. A: Math. Theor., 40(26):7255–7264, June 2007.
- [93] **P. Jordan, J. von Neumann, and E. Wigner.** *On an Algebraic Generalization of the Quantum Mechanical Formalism*. Annals of Mathematics, 35(1):29–64, 1934.
- [94] **A. A. Albert.** *On Jordan algebras of linear transformations*. Transactions of the American Mathematical Society, 59(3):524–524, March 1946.
- [95] **Y. Aharonov and D. Bohm.** *Significance of Electromagnetic Potentials in the Quantum Theory*. Physical Review, 115(3):485–491, August 1959.
- [96] **V. Vedral.** *Geometric Phases and Topological Quantum Computation*. arXiv:quant-ph/0212133, December 2002.
- [97] **L. E. J. Brouwer.** *Über Abbildung von Mannigfaltigkeiten*. Mathematische Annalen, 71(1):97–115, March 1911.
- [98] **R. P. Agarwal, M. Meehan, and D. O’Regan.** *Fixed Point Theory and Applications*. Cambridge University Press, Cambridge, 2001.
- [99] **F. R. Gantmacher.** *The theory of matrices, Volume 2*. AMS Chelsea Publishing, 1959.

- [100] **T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski.** *On mutually unbiased bases.* International Journal of Quantum Information, 8(4):535–640, April 2010.
- [101] **O. C. O. Dahlsten.** *Quantum theory escapes locality by accepting uncertainty,* FQXi Community Blog Post, August 2014. <http://fqxi.org/community/forum/topic/2233>.
- [102] **D. Bohm.** *A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I.* Physical Review, 85(2):166–179, January 1952.
- [103] **D. Bohm.** *A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. II.* Physical Review, 85(2):180–193, January 1952.
- [104] **D. Dürr and S. Teufel.** *Bohmian Mechanics: The Physics and Mathematics of Quantum Theory.* Springer-Verlag, Berlin, 2009.
- [105] **X. Oriols and J. Mompart.** *Applied Bohmian Mechanics.* Pan Stanford Publishing, June 2012.
- [106] **T. Norsen.** *The pilot-wave perspective on spin.* American Journal of Physics, 82(4):337–348, April 2014.
- [107] **V. I. Arnold.** *Mathematical Methods of Classical Mechanics.* Springer-Verlag, New York, second edition, 1978.
- [108] **A. Einstein.** *Zur Elektrodynamik bewegter Körper.* Annalen der Physik, 14(S1):194–224, February 1905.
- [109] **J. B. Fraleigh.** *A First Course in Abstract Algebra.* Pearson, Harlow, seventh edition, 2002.
- [110] **M. R. Sepanski.** *Complex Lie groups.* Springer, New York, 2007.
- [111] **E. M. Alfsen and F. W. Shultz.** *Geometry of State Spaces of Operator Algebras.* Birkhäuser, Boston, 2003.
- [112] **M. A. Graydon.** *Quaternions and Quantum Theory.* Master's thesis, University of Waterloo, 2011.
- [113] **S. L. Adler.** *Quaternionic Quantum Mechanics and Quantum Fields.* Oxford University Press, Oxford, 1995.
- [114] **Ll. Masanes, M. P. Müller, D. Perez-Garcia, and R. Augusiak.** *Entanglement and the three-dimensionality of the Bloch ball.* arXiv:1111.4060, November 2011.

- [115] **J. C. Baez.** *Symplectic, Quaternionic, Fermionic*, July 2014. <http://www.math.ucr.edu/home/baez/symplectic.html>.
- [116] **B. Dakić and Č. Brukner.** *The classical limit of a physical theory and the dimensionality of space*. arXiv:1307.3984, July 2013.
- [117] **U. Sinha, C. Couteau, T. Jennewein, R. Laflamme, and G. Weihs.** *Ruling out multi-order interference in quantum mechanics*. Science (New York, N.Y.), 329(5990):418–21, July 2010.
- [118] **I. Söllner, B. Gschösser, P. Mai, B. Pressl, Z. Vörös, and G. Weihs.** *Testing Born’s Rule in Quantum Mechanics for Three Mutually Exclusive Events*. Foundations of Physics, 42(6):742–751, September 2011.
- [119] **G. Weihs.** *Higher-Order Interferences of Single Photons*. Presentation at ICQIQ (Bangalore, India), January 2013.
- [120] **A. Peres.** *Proposed Test for Complex versus Quaternion Quantum Theory*. Physical Review Letters, 42(11):683–686, March 1979.
- [121] **M. Massimi.** *Pauli’s Exclusion Principle: The Origin and Validation of a Scientific Principle*. Cambridge University Press, Cambridge, 2005.
- [122] **S. N. Bose.** *Plancks Gesetz und Lichtquantenhypothese*. Zeitschrift für Physik, 26(1):178–181, December 1924.
- [123] **F. Dalfovo, S. Giorgini, and S. Stringari.** *Theory of Bose-Einstein condensation in trapped gases*. Reviews of Modern Physics, 71(3):463–512, April 1999.
- [124] **L. Pitaevskii and S. Stringari.** *Bose-Einstein Condensation*. Oxford University Press, Oxford, 2003.
- [125] **J. M. Leinaas and J. Myrheim.** *On the theory of identical particles*. Il Nuovo Cimento B, 37(1):1–23, January 1977.
- [126] **F. Wilczek.** *Quantum Mechanics of Fractional-Spin Particles*. Physical Review Letters, 49(14):957–959, October 1982.
- [127] **W. Pauli.** *The Connection Between Spin and Statistics*. Physical Review, 58: 716–722, 1940.
- [128] **M. Fierz.** *Über die relativistische Theorie kräftefreier Teilchen mit beliebigem Spin*. Helv. Phys. Acta, 12(1):3–37, 1939.
- [129] **L. Hardy.** *Foliable Operational Structures for General Probabilistic Theories*. arXiv:0912.4740, December 2009.

- [130] **C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma.** *Non-Abelian anyons and topological quantum computation.* Reviews of Modern Physics, 80(3):1083–1159, September 2008.
- [131] **J. K. Pachos.** *Introduction to topological quantum computation.* Cambridge University Press, Cambridge, 2012.
- [132] **D. Oriti** (editor). *Approaches to quantum gravity.* Cambridge University Press, Cambridge, 2009.
- [133] **R. P. Feynman and A. R. Hibbs.** *Quantum Mechanics and Path Integrals.* Dover, New York, emended edition, 1965.