Bounds on computation from physical principles

Ciarán M. Lee

University College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2016

For Maisie D'Arcy.
It's all double-dutch to me too.

I want to be
Like the man
Who drops at night
From a moving train

And strikes out over the fields
Where fireflies glow
Not knowing a word of the language.

Excerpt from “The Last of the Fire Kings”
Derek Mahon

Acknowledgements

Many people need to be thanked for getting me through this thesis and (more importantly?) keeping me sane throughout the process. First and foremost I am indebted to my supervisor, Jon Barrett. While he may not be a vampire, he is most certainly a gentleman—penchant for lobster notwithstanding. Jon has an uncanny ability to make even the most complicated topics clear and understandable. A feat I have tried to imitate many times without much success. Working with Jon has been an honour. I am also indebted to Rob Spekkens for encouraging my interest in foundational questions during my masters studies at PI, and inviting me back to visit during my doctoral studies. I couldn't have asked for better teachers than Jon and Rob. I would like to thank my examiners, Richard Jozsa and Vlatko Vedral, for their careful reading of this thesis. Any remaining inaccuracies are of course my own. I also want to thank Bob Coecke, Samson Abramsky, and Jamie Vicary for all their help.

Throughout my doctorate I had the good fortune to learn from three great post-docs: Matty Hoban, Niel de Beaudrap, and Nathan Walk. All three of them ensured the Quantum Group was a great place to talk science. More importantly, they made it a fun place to be. Fairplay to the lads. I especially want to thank John Selby for kindly using his mad tikZit skilz to draw most of the figures in this thesis, and for his monthly pilgrimages out beyond the ring road in search of Barry's tea. I couldn't ask for a better collaborator, or friend. Some man. I also want to thank John-Mark, Carlo-Maria, and the rest of my office mates. Top lads.

I want thank my parents , Joanne and Damien, and my two sisters, Aisling and Fiona, for all their love and support. This thesis is for them. Finally, and most importantly, I want to thank Katie Gilligan for her love, intelligence, and for just being a joy to be around. There's a reason you should go out with your best friend. Kates, you kept me sane throughout the last three years and provided me with perspective when I needed it most. I couldn't have done this without you.

Abstract

The advent of quantum computing has challenged classical conceptions of which problems are efficiently solvable in our physical world. This raises the general question of what broad relationships exist between physical principles and computation. The current thesis explores this question within the operationally-defined framework of generalised probabilistic theories. In particular, we investigate the limits on computational power imposed by simple physical principles. At present, the best known upper bound on the power of quantum computers is that $\mathbf{BQP} \subseteq \mathbf{AWPP}$, where **AWPP** is a classical complexity class contained in **PP**. We define a circuit-based model of computation in the above mentioned operational framework and show that in theories where local measurements suffice for tomography, efficient computations are also contained in **AWPP**. Moreover, we explicitly construct a theory in which the class of efficiently solvable problems exactly equals **AWPP**, showing this containment to be tight. We also investigate how simple physical principles bound the power of computational paradigms which combine computation and communication in a non-trivial fashion, such as interactive proof systems. Additionally, we show how some of the essential components of computational algorithms arise from certain natural physical principles. We use these results to investigate the relationship between interference behaviour and computational power, demonstrating that non-trivial interference behaviour is a general resource for post-classical computation. We then investigate whether post-quantum interference is a resource for post-quantum computation. Sorkin has defined a hierarchy of possible post-quantum interference behaviours where, informally, the order in the hierarchy corresponds to the number of paths that have an irreducible interaction in a multi-slit experiment. In quantum theory, at most pairs of paths can ever interact in a fundamental way. We consider how Grover's speed-up depends on the order of interference in a theory, and show that, surprisingly, the quadratic lower bound holds regardless of the order of interference.

Publications

The majority of the work in this thesis is based on the following publications:

C. M. Lee and J. Barrett, *Computation in generalised probabilistic theories*, New Journal of Physics 17, 083001, 2015.

C. M. Lee and M. J. Hoban, *Bounds on the power of proofs and advice in general physical theories*, Proc. R. Soc. A 472 (2190), 20160076 (2016).

C. M. Lee and M. J. Hoban, *The Information Content of Systems in General Physical Theories*, EPTCS 214, 2016, pp. 22-28, 2016.

C. M. Lee and J. H. Selby, *Higher-order interference in extensions of quantum theory,* Foundations of Physics, 2016, DOI:10.1007/s10701-016-0045-4.

C. M. Lee and J. H. Selby, *Generalised phase kick-back: the structure of computational algorithms from physical principles,* New J. Phys. 18 (2016) 033023.

C. M. Lee and J. H. Selby, *Deriving Grover's lower bound from simple physical principles,* New J. Phys. 18 (2016) 093047.

Contents

Introduction **1**

1 Generalised probabilistic theories **7**

1.1 Devices and circuits 9

1.2 Probabilistic structure 12

1.3 Physical principles 18

1.3.1 Tomographic locality 18

1.3.2 Causality 21

1.3.3 Purification and purity preservation 22

1.3.4 Strong symmetry 24

1.4 Examples 25

1.4.1 Quantum theory 25

1.4.2 Classical probability theory 27

1.4.3 Real Hilbert space quantum theory 28

1.4.4 PR boxes, Boxworld, and polygon theories 29

1.4.5 A non-causal theory 31

2 Computation in generalised probabilistic theories **33**

2.1 The computational model 35

2.1.1 Uniform circuits 35

2.1.2 Acceptance criterion 37

2.1.3 Efficient computation 38

2.2 Upper bounds on computational power 39

2.3 Post-selection and generalised probabilisitic theories 42

2.4 Oracles 44

2.5 Proofs of the theorems 47

2.5.1 Approximate circuit families 47

2.5.2 Proof of Theorem 2.2 50

2.5.3 Proof of Theorem 2.7 54
2.5.4 Proof of Theorem 2.10 55
2.6 Discussion and conclusion 57

3 The computational landscape of generalised theories 59
3.1 The modified framework 61
3.2 Evidence against free theories achieving the upper bound 64
3.3 Achieving the upper bound with a non-free theory 66
3.4 Proof of theorem 3.2 67
3.4.1 Affine Turing Machines 67
3.4.2 Solving **AWPP** problems with an affine Turing machine . . . 68
3.4.3 Simulating an Affine Turing Machine in **AWPP** 69
3.4.4 Constructing affine circuits 70
3.4.5 A tomographically local theory 73
3.5 Discussion 78

4 Bounding the power of proofs and advice 82
4.0.1 Overview of the results 83
4.1 Proofs and advice 85
4.1.1 Definitions for general theories 86
4.1.2 Example: Boxworld 88
4.2 Consequences of non-trivial dynamics for computation 89
4.2.1 Powerful computation from non-trivial dynamics 89
4.2.2 Bounds on computation with advice in physical theories . . . 90
4.3 Bounds on the power of proofs in physical theories 94
4.4 Relating proofs and advice? 97
4.4.1 Related work 98
4.5 Discussion and conclusion 99

5 Higher-order interference 101
5.1 Background and introduction 102
5.2 A definition of higher-order interference in generalised probabilistic theories 106
5.3 Density cubes 109
5.3.1 States and effects 109
5.3.2 Transformations 111
5.3.3 Hyper-decoherence 112

5.3.4 A computational advantage? 115
5.3.5 Issues with the Density Cube framework 116
5.3.5.1 Axioms insufficient to specify a unique operational theory 116
5.3.5.2 Characterising the set of physical transformations . 118
5.4 Quartic quantum theory 119
5.4.1 Description of the theory 120
5.4.2 Interference in Quartic Quantum Theory 121
5.4.3 Composite systems in Quartic Quantum Theory 123
5.4.3.1 Note on Boxworld-like correlations in Quartic Quantum Theory 124
5.5 Conclusion 124

6 Generalised phase kick-back: the physical structure of algorithms 126
6.1 Higher-order interference via phase transformations 127
6.1.1 A quantum example 127
6.1.2 Operational theories 129
6.1.3 Higher-order interference and phase 131
6.2 Controlled transformations and a generalised phase kick-back 133
6.2.1 Particle exchange experiments 135
6.2.2 Computational oracles 137
6.3 General results following from physical principles 2 to 5 140
6.3.1 Self-duality and uniqueness of distinguishing measurements . . 141
6.3.2 Existence of a completely mixed state 141
6.3.3 Purification of the completely mixed state is dynamically faithful141
6.4 Proof of theorem 6.12 142
6.5 Proof of lemma 6.13 145
6.6 Proof of lemma 6.14 146
6.7 Proof of theorem 6.16 147
6.8 Conclusion 149

7 Higher-order interference doesn't help in searching for a needle in a haystack[*] 151
7.1 Standing assumptions 152
7.2 Higher-order interference in the presence of principles 2 to 5 153

[*] With apologies to L. Grover [149]

7.3 Setting up the problem 157
7.4 Main result 159
7.5 Discussion 162

Summary and future work **163**

Appendix A A compendium of complexity classes **169**

Appendix B Proofs and advice: the classical case **171**

Appendix C Proofs and advice: the quantum case **173**

Appendix D Comparing definitions of higher-order interference **174**

Bibliography **176**

List of Figures

1.1 Connections between certain physical devices 8
1.2 State space of a single Boxworld system 30

3.1 Simulation of an affine Turing machine by an affine circuit 67
3.2 Depiction of sub-circuit B. 72
3.3 The lefthand side schematically depicts the set of non-signalling theories, ranked in increasing order (from the center of the figure) of Bell inequlaity violation. The righthand side schematically depicts the set of all tomographically local theories, ranked in increasing order (from bottom to top) of the power of their efficient computation. 80

5.1 Depiction of a multi-slit experiment 106

6.1 Relation between the physical set-up of Mach-Zender interferometer and the operational description, depicted in circuit notation 128
6.2 Schematic illustration showing a particle exchange experiment involving two particles. 136
6.3 Schematic illustration showing a particle exchange experiment involving mulitple particles. 137

7.1 Querying an oracle using a multi-slit experiment 158

A.1 Relation between certain computational complexity classes. A bold line between two classes indicates the lower class is contained within the higher. 170

Introduction

Since the discovery of quantum theory in the early twentieth century, its predictions have been an affront to our classically honed sensibilities. From superposition and interference to entanglement, steering, and non-locality, quantum theory boasts a litany of classically counter-intuitive features. Many physicists have found these features hard to stomach, as demonstrated by Einstein's infamous dismissal of steering as "spooky action-at-a-distance[1]" [2] and Feynman's description [3] of interference as "impossible, *absolutely* impossible, to explain in any classical way". However, the advent of quantum information science in the 1980's brought with it a major conceptual breakthrough which caused a radical shift in physicists perceptions of these seemingly paradoxical features: Quantum theory appears to offer dramatic advantages [1, 4] for various information-processing tasks—computation in particular [1, 4, 38, 64, 63, 80, 81, 82, 83, 84, 134, 149, 150].

The realisation that interference [4, 84] and entanglement [81] are resources for post-classical information processing provides a novel understanding of these seemingly paradoxical quantum features. This raises the general question of what broad relationships exist between natural physical principles[2] and computation. For instance, what limits on computation are imposed by physical principles? Can the general structure of computational algorithms be seen to arise from physical principles alone? Understanding what general relationships hold may further deepen our understanding of quantum theory. Moreover, systematically investigating such relationships could illuminate the infamous [85, 86, 87] source of the quantum computational "speed-up".

Such questions are similar in spirit to recent attempts at characterising the set of quantum correlations solely from simple physical principles, such as information causality [77] and local orthogonality [76]. These principles, while not fully capturing

[1] Or "spukhafte fernwirkung" in the original German [2].

[2] An example would be the no-signalling principle. Later in this section we informally introduce the principles we will be studying in this thesis, with the formal defintions provided in chapter 1.

the exact quantum boundary [75], have deepened our understanding of quantum correlations and have even led to the development of Device-Independent Cryptography [10]. Hence, while investigating such connections has foundational interest, it has also been shown to have practical implications.

In recent years, one approach to uncovering the origin of the quantum "speed-up" has been to ask how computational power changes as features of quantum theory are altered. Beginning with the work of Abrams and Lloyd, it was shown that supplimenting quantum theory with exotic transformations can result in the trivial solution to computationally hard problems [12]. Unfortunately, changing various aspects of quantum theory in a ad-hoc manner does not, in general, lead to a consistent physical theory. This has even motivated the belief that quantum theory is an "island" within the space of all possible theories; alter quantum mechanics and we obtain dramatic consequences [65]. We thus require an abstract framework in which to study the power of computation, where quantum and classical computation are special cases.

In this thesis we investigate computation in the operational-defined framework of generalised probabilistic theories [15, 17, 19]. Informally, a theory in this framework specifies a set of laboratory devices that can be connected together in different ways and assigns probabilities to different experimental outcomes. Theories within this framework can be described that are different from classical or quantum theory, but which nonetheless make good operational sense and allow one to systematically investigate the connections between physical principles and information-theoretic advantages. This framework suggests a natural model of computation, analogous to the classical and quantum circuit models, which we shall use to rigorously investigate the connections between physical principles and computational power. From bounding the class of problems a general theory can efficiently solve to deriving the general structure of computational algorithms from physical principles, we shall explore the multifaceted relationship between computation and physical principles in this framework from many different perspectives.

In chapter 1 we introduce the framework of generalised probabilistic theories and discuss five natural physical principles which we shall connect to computation over the course of this thesis: causality (roughly, probabilities for present experiments cannot depend on future measurement choices), tomographic locality (local measurements suffice for tomography), purification (information can never be destroyed, only discarded), purity preservation (composition preserves information) and strong symmetry (the existence of non-trivial reversible dynamics). We also provide examples of concrete theories, distinct from classical and quantum theory, which illustrate the fact that the above physical principles are logically independent; generalised probabilistic theories can be specified that satisfy any subset of the five principles above[3]. Moreover, the existence of such alternate theories allows for a rigorous investigation of the structural or information-theoretic consequences of different physical principles.

In chapter 2, we introduce a circuit-based model of computation in the generalised probabilistic theory framework and use it to rigorously define the class of problems a specific theory can solve efficiently. With this definition in hand, we begin to explore the limits on efficient computation imposed by certain physical principles. At present, the best upper bound known for the power of quantum computation is that **BQP** $\subseteq$ **AWPP**, where **AWPP** is a classical complexity class known to be included in **PP**, hence **PSPACE**. This was first proved by Fortnow and Rogers in [45]. See appendix A for a review of all computational complexity classes discussed in this thesis. What is the minimal set of physical principles under which the above inclusion holds for efficient computation in a specific generalised probabilistic theory? We show that given only an assumption of tomographic locality, efficient computations are contained in **AWPP**. This inclusion still holds even without assuming the principle of causality. Following Aaronson, we extend the computational model by allowing post-selection of measurement outcomes. Aaronson showed that the corresponding quantum complexity class, **PostBQP**, is equal to **PP**. Given only the assumption of tomographic locality, the inclusion in **PP** still holds for post-selected computation in general theories. Hence in a world with post-selection, quantum theory is optimal for computation in the space of all operational theories. We also consider whether one can define computational oracles and obtain relativised complexity results for general theories. The results presented in this chapter first appeared in [93] and constitute joint work with J. Barrett.

In chapter 3 we investigate whether the bound derived in chapter 2 on efficient computation in tomographically local theories is tight. We provide complexity-theoretic arguments which suggest that this is unlikely. However, by slightly modifying[4] the definition of what constitutes a generalised probabilistic theory, one can construct a theory within this altered framework—satisfying both tomographic locality and causality—in which the class of efficiently solvable problems exactly equals **AWPP**. Hence **AWPP**, despite having a slightly involved definition in terms of "gap functions" for non-deterministic Turing machines (see chapter 2), can be thought of

[3] With the exception of causality and purification, see chapter 1.

[4] Instead of assigning probabilities to any experiment composed of laboratory devices, a theory in the modified framework only assigns probabilities to certain *allowed* experiments, see chapter 3 for a rigorous discussion.

much more intuitively as the class of problems efficiently solvable by tomographically local physical theories. The theory constructed in this chapter has the maximum computational power consistent with tomographic locality. In a sense, one can think of it as the analogue of a PR-box—which exhibits the strongest non-local correlations consistent with the no-signalling principle—for computation. The main result of this chapter is joint work with J. Barrett, N. de Beaudrap, and M. J. Hoban and will appear in [94].

In chapter 4 we investigate how simple physical principles bound the power of two different computational paradigms which combine computation and communication in a non-trivial fashion: computation with advice and interactive proof systems. We show that the existence of non-trivial dynamics in a theory, which is guaranteed if the theory satisfies the principle of strong symmetry, implies a bound on the power of computation with advice. Moreover, we provide an explicit example of a theory with no non-trivial dynamics in which the power of computation with advice is unbounded. Finally, as was the case for efficient computation in chapter 2, we show that the power of simple interactive proof systems in theories satisfying tomographic locality satisfies the best bound known in the quantum case. These results first appeared in [66, 67] and are joint work with M. J. Hoban.

As we have seen over the last few paragraphs, chapters 2 to 4 use the language of complexity classes to derive general bounds on the power of computation in generalised probabilistic theories. However, much of quantum computing is concerned not so much with the high-level view offered by complexity classes, but instead with the construction of concrete algorithms to solve specific problems. Over the course of chapters 5, 6, and 7, we shall investigate both the structure of computational algorithms in generalised probabilistic theories and whether certain algorithmic advantages are directly related to physical principles.

Quantum interference between computational paths has been posited [84] as a key resource behind the computational "speed-ups" offered by many quantum algorithms, such as Grover's search algorithm [149]. However, as first noted by Rafael Sorkin [124, 125], there is a limit to quantum interference—at most pairs of paths can ever interact in a fundamental way. Sorkin has defined a hierarchy of possible 'higher-order' interference behaviours—currently under experimental investigation [131, 132, 153]—where classical theory is at the first level of the hierarchy and quantum theory belongs to the second. Informally, the order in the hierarchy corresponds to the number of paths that have an irreducible interaction in a multi-slit experiment. Could more interference imply more computational power?

Chapter 5 provides an overview of the literature on higher-order interference. It also investigates two proposed extensions of quantum theory from the point of view of their interference behaviour: the theory of Density Cubes proposed by Dakić, Paterek and Brukner, which has been shown to exhibit third-order interference in a three slit experiment, and the Quartic Quantum Theory of Życzkowski. This investigation clarifies the impact of these two generalised theories to ongoing experimental tests for higher-order interference, and explores potential information-theoretic consequences of post-quantum interference in this concrete setting. This investigation first appeared in [88] and is joint work with J. H. Selby.

In chapter 6 we show that some of the essential machinery of quantum computation—namely reversible controlled transformations and the phase kick-back mechanism—exist in any operationally-defined theory which satisfies the principles of causality, purification, purity preservation, and strong symmetry. We use these results to investigate the relationship between interference behaviour and computational power, demonstrating that non-trivial interference behaviour is a general resource for post-classical computation. In proving the above, we connect post-quantum interference to the existence of post-quantum particle types, potentially providing a novel experimental test for higher-order interference. We also introduce a framework that relates higher-order interference to *phase transformations* in operationally-defined theories. Finally, we show that reversible controlled transformations in theories satisfying the above principles give rise to computational oracles which solve the subroutine problem of Bennett et al. from [44]. That is, in theories satisfying causality, purification, purity preservation, and strong symmetry we have that $\mathbf{BGP}^{\mathbf{BGP}} = \mathbf{BGP}$. The results in this chapter first appeared in [89] and is joint work with J. H. Selby.

In chapter 7 we consider how Grover's speed-up depends on the order of interference in a generalised theory. We consider theories satisfying causality, purification, purity preservation, and strong symmetry, which, as we mentioned in the previous paragraph, are sufficient for the existence of reversible controlled transformations and hence well-defined search oracles. Given these principles, we prove that a theory at level h in Sorkin's hierarchy requires $\Omega(\sqrt{N/h})$ queries to solve the search problem. Thus, at least from the point of view of the search problem, post-quantum interference does not imply a computational speed-up over quantum theory. Moreover, one can view this result as a derivation of the quadratic lower bound to the search problem from physical principles, the computational analogue of deriving Cirel'son's bound [6] from simple physical principles. This result constitutes joint work with J. H. Selby and first appeared in [90].

Finally, in the summary and further work section, we discuss possible practical implications of the above results. In particular, we argue that these results lay the groundwork for developing protocols to verify delegated computation which are secure against adversaries with the ability to perform post-quantum dynamics. This bears a strong resemblence to the construction of quantum key distribution protocols which are secure against adversaries with the ability to generate post-quantum correlations [10]. Such protocols would provide a way to verify delegated computations whose correctness is derived directly from simple physical principles, similar to cryptographic schemes whose security rests solely on the no-signalling principle [10].

Chapter 1

Generalised probabilistic theories

Quantum theory is a strange beast; its predictions have been verified to unprecedented accuracy, yet the standard quantum formalism—in which quantum states are represented by positive operators[1] acting on an underlying complex Hilbert space—is as abstract as its predictions are accurate. Despite being universally accepted among physicists as a tool for calculating the probabilities of possible experimental outcomes, the standard language of complex Hilbert spaces lacks direct physical or operational significance. As Asher Peres [24] famously put it: "Quantum phenomena do not occur in a Hilbert space. They occur in a laboratory".

The history of physics is replete with similar examples. For instance, when the Lorentz transformations were first discovered by FitzGerald and Lorentz [25], they were somewhat ad hoc and rather abstract. However, they were soon seen to arise in a natural manner from Einstein's two principles of special relativity: the laws of physics are invariant in all inertial reference frames, and the speed of light in a vacuum is a constant for all observers, regardless of the source. Before he could state his principles, Einstein first had to introduce an operational framework, based on clocks and rods, in order to define the notion of a reference frame. Quantum theory is ad hoc and abstract in the same manner as the Lorentz transformations were before Einstein and special relativity. An approach to better understanding quantum theory then suggests itself: can an operational framework be devised—akin to Einstein's clocks and rods—from which quantum theory could be seen to naturally arise from simple physical principles? The formalism of generalised probabilistic theories [15, 16, 18, 19], which has been gaining much interest in recent years, provides such a framework.

The fundamental goal of any physical theory is to provide a consistent account of experimental data. This constitutes the core idea underlying the framework of gener-

[1] Or, more accurately, positive semi-definite operators

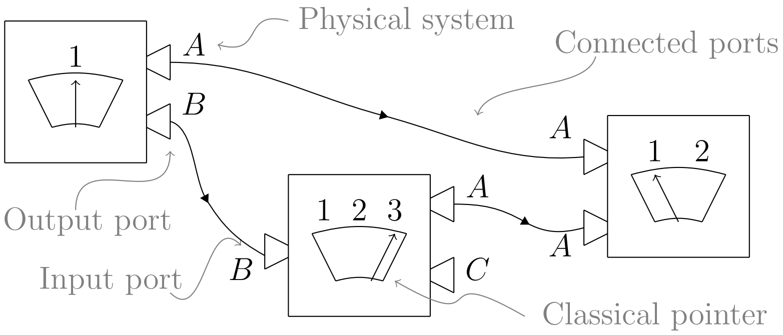


Fig. 1.1: Connections between certain physical devices

alised probabilistic theories, where the primitive notions are operational[2] in nature. Informally, a theory in this framework specifies a set of laboratory devices that can be connected together in different ways, as schematically illustrated in Fig. 1.1, and assigns probabilities to different experimental outcomes. Using this framework, many authors have derived the entire structure of quantum theory from simple physical principles [16, 18, 19, 20, 21, 22, 72] in a manner akin to Einstein's derivation of the Lorentz transformations from two easily stated physical principles.

A remarkable feature of this operational framework is that it provides examples of theories which differ from quantum theory, yet still make good operational sense. An obvious example is classical probability theory, which can be used to calculate probabilities for classicial situations such as tossing a coin, or conducting an experiemnt in the regime of Newtonian physics. More exotic examples include Spekkens' toy theory [123, 126] and a construction colloquially known as "Boxworld" [17, 29], which achieves the largest possible violation of the CHSH inequality [71] consistent with the no-signalling principle.

The existence of such alternate theories allows for an investigation of the structural, or information-theoretic, properties of theories where different physical principles may hold. Such an investigation could provide a deep understanding of the connections between physical principles and information-theoretic advantages in a theory-independent manner. For instance, instead of asking such questions as "does quantum interference imply an advantage for quantum computation", one can now ask: "does the existence of interference in any physical theory imply post-classical computation?". While models such as the toy theory and boxworld might not corre-

[2] Note that operationalism as a philosophical viewpoint, in which one asserts that there is no reality beyond laboratory device settings and outcomes, is not being espoused here. One should merely view the approach taken here as an operational methodology aimed at gaining insight into certain structural properties of physical theories.

spond to descriptions of our physical world, they nevertheless make good operational sense and allow one to systematically investigate the connections between physical principles and information-theoretic advantages [17, 15].

Much progress has already been made in understanding the connections between physical principles and some information-theoretic tasks, such as cryptography and communication complexity problems. It is now known that the degree of non-locality in a theory is related to its ability to solve communication complexity problems [5] and to its ability to perform super-dense coding, teleportation and entanglement swapping [7]. Teleportation and no-broadcasting are now better understood than they were when investigated solely from the viewpoint of quantum theory [8, 9, 15]. Moreover, cryptographic protocols have been developed whose security relies not on aspects of the quantum formalism, but on general physical principles. The above forms part of the broader research program of generalised probabilistic theories, which, in the words of Barnum, Müller, and Ududec [72], aims to "analyse the structure of physics—that is, the way that the different parts of physics fit together—by rigorously assessing the consequences of changing some of its parts".

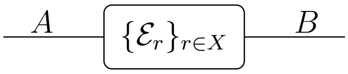
We will work in the circuit framework for generalised probabilistic theories developed by Hardy in [19] and Chiribella, D'Ariano, and Perinotti in [15, 16]. The circuit framework takes inspiration from the categorical approach to quantum theoy, introduced by Abramsky and Coecke [141, 140], in that it heavily emphasises compositionality. In this manner it is different to the convex sets approach to generalised probabilistic theories presented in [17, 18]—although both are similar in spirit. The presentation of generalised probabilisitc theories given in this chapter is most similar to that of Chiribella et al., see [23] for another review of the framework.

The basic framework of generalised probabilistic theories will be presented over the course of section 1.1 and 1.2. Section 1.3 provides rigorous definitions of the physical principles we will be concerned with in this thesis. Finally, section 1.4 provides detailed examples of theories in this framework.

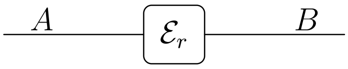
1.1 Devices and circuits

The idea of a generalised probabilistic theory is that a set of physical, or laboratory, devices is specified, which can be connected together in different ways, such that the theory assigns probabilities for different outcomes of said devices. As depicted in Fig. 1.1, each physical device comes attached with input ports, output ports, and a classical pointer. Whenever a device is used in an experiment the classical pointer ends up in one of a number of positions, indicating a specific outcome has occurred. Input and output ports are typed, with types given by labels $A, B, C \ldots$. As discussed in more detail below, physical devices can be composed both sequentially and in parallel, and when composed sequentially, types must match: the output ports of the first device must have the same types as the corresponding input ports of the second.

Suppose that for a particular device, the classical outcome r takes values in a set X. We shall assume throughout that $|X|$ is finite. A device $\mathcal{E}$, with specified input and output types, then defines a set of *events*, one for each classical outcome, $\{\mathcal{E}_r\}_{r\in X}$. With an input port of type A and an output port of type B, for example, the device can be represented diagrammatically as



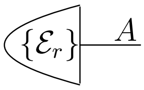
and a specific event as



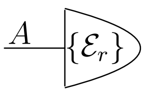
A device is *deterministic* if its outcome set X is the singleton set.

Although devices, with input and output ports, and a pointer, form the primitives of the operational theory, it is also useful to introduce a notion of *physical system.* A system may be thought of as passing between the output port of a device, and the input port of the next, and has the same type as the ports. In other words, in the diagrams above and below, systems correspond to wires. Given two systems of types A and B, we can form a *composite system* of type AB. Operationally, a device with input system AB corresponds to a physical device with a set of input ports labelled by A and a disjoint set of input ports labelled by B.

A device with no input ports corresponds to preparing a system—more precisely, such a device corresponds to a set of preparations, with the classical pointer indexing which preparation actually occurs. Such a device can be represented diagrammatically as:



A device with no output ports corresponds to a measurement (that destroys or discards the system), with the classical pointer indexing the measurement outcome. Diagrammatically, such a device can be written:

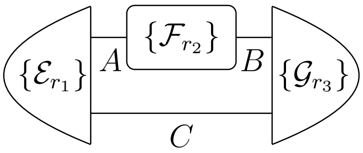


Both device and events can be composed in sequence and in parallel. If $\{\mathcal{E}_{r_1}\}_{r_1\in X_1}$ is a device from system A to B and $\{\mathcal{U}_{r_2}\}_{r_2\in X_2}$ is a device from system B to C, then their sequential composition is a device from A to C with outcomes $(r_1, r_2) \in X_1 \times X_2$ and events $\{\mathcal{U}_{r_2} \circ \mathcal{E}_{r_1}\}_{(r_1,r_2)\in X_1\times X_2}$. Similarly, if $\{\mathcal{E}_{r_1}\}_{r_1\in X_1}$ is a device from system A to B and $\{\mathcal{U}_{r_2}\}_{r_2\in X_2}$ is a device from system C to D, then their parallel composition is a device from the composite system AC to the composite system BD with outcomes $(r_1, r_2) \in X_1 \times X_2$ and events $\{\mathcal{U}_{r_2} \otimes \mathcal{E}_{r_1}\}_{(r_1,r_2)\in X_1\times X_2}$. Note that the symbol $\otimes$—which schematically denotes parallel composition—may not correspond to the standard vector space tensor product[3]. Sequential and parallel composition satisfy

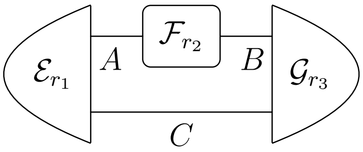
$$\left(\mathcal{U}_{r_3} \otimes \mathcal{E}_{r_4}\right) \circ \left(\mathcal{F}_{r_1} \otimes \mathcal{K}_{r_2}\right) = \left(\mathcal{U}_{r_3} \circ \mathcal{F}_{r_1}\right) \otimes \left(\mathcal{E}_{r_4} \circ \mathcal{K}_{r_2}\right)$$

for every $\mathcal{U}_{r_3}, \mathcal{E}_{r_4}, \mathcal{F}_{r_1}, \mathcal{K}_{r_2}$ with the property that the output of $\mathcal{F}_{r_1}$ (respectively, $\mathcal{K}_{r_2}$) matches the input of $\mathcal{U}_{r_3}$ (respectively, $\mathcal{E}_{r_4}$). A generalised probabilistic theory specifies a set of devices, closed under sequential and parallel composition.

A circuit in a generalised probabilistic theory corresponds to a number of devices, connected in sequence and in parallel, such that there are no unconnected ports (i.e., no dangling input or output wires), and no cycles. For example:



A specific outcome of the above circuit corresponds to a particular classical outcome for each of the tests, i.e., to a collection of events, connected in sequence and in parallel:



(1.1)

In the following, we shall use the term *circuit fragment* to denote a number of devices, connected in sequence and in parallel, such that there are some unconnected ports but adding in a device with ports whose type matches the unconnected ports results in a closed circuit.

[3] In theories satisfying tomographic locality, which shall be defined in section 1.3.1, the symbol $\otimes$ does indeed correspond to the standard vector space tensor product.

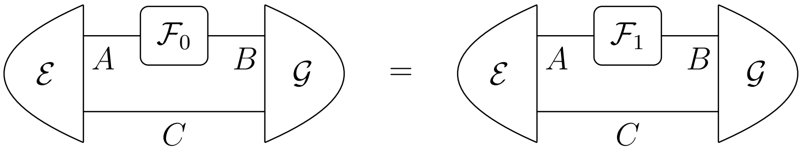
1.2 Probabilistic structure

So far we have described the operational part of a generalised probabilistic theory, but not the probabilistic part. In addition to specifying a set of devices, hence sets of circuits and circuit outcomes, a probabilistic theory should assign probabilities to circuit outcomes. In a generalised probabilistic theory, every outcome of a circuit is assigned a probability $P(r_1r_2\ldots r_n)$, understood as the joint probability of outcomes $r_1,\ldots,r_n$ for the individual devices occurring on a single run. The joint probabilities satisfy $\sum_{r_1r_2\ldots r_n} P(r_1r_2\ldots r_n) = 1$. A further constraint is that probabilities for unconnected, i.e. independent, circuits factorise. This means that for events $\mathcal{E}_{r_1r_2\ldots r_m}$ and $\mathcal{F}_{s_1s_2\ldots s_n}$, each of which corresponds to the outcome of a closed circuit, probabilities assigned to the composite events $\mathcal{E}_{r_1r_2\ldots r_m}\otimes\mathcal{F}_{s_1s_2\ldots s_n}$, and $\mathcal{E}_{r_1r_2\ldots r_m}\circ\mathcal{F}_{s_1s_2\ldots s_n}$, each satisfy $P(r_1\ldots r_m, s_1\ldots s_n) = P(r_1\ldots r_m)P(s_1\ldots s_n)$. The operational content of this assertion is that the probabilities of the outcomes in any circuit depend only on the tests and connections given, and not on external tests and events[4].

We can now formally define a generalised probabilistic theory.

Definition 1.1 (Generalised probabilistic theory)**.** *A generalised probabilistic theory* $\boldsymbol{G}$ *is specified by a collection of devices which are closed under parallel and sequential composition, such that each closed circuit corresponds to a collection* $\{p_i\}_{i\in Z}$ *where* $p_i\in[0,1]$, $\sum_i p_i = 1$, *and* Z *is the set of all outcomes of the closed circuit. Moreover, probabilities for independent circuits factorise.*

The introduction of probabilities into the theory induces linear structure that will be crucial in what follows. Consider two events $\mathcal{F}_0$ and $\mathcal{F}_1$, whose input and output ports have matching types. Suppose that for every closed circuit, and every outcome of the circuit, replacing $\mathcal{F}_0$ with $\mathcal{F}_1$ does not change the probability of the outcome. In this case, $\mathcal{F}_0$ and $\mathcal{F}_1$ are *equivalent.* Pictorially, if



for all preparation events $\mathcal{E}$ of system AC, measurement events of system BC, and for all systems C, then event $\mathcal{F}_0$ is equivalent to event $\mathcal{F}_1$. The events $\mathcal{F}_0$ and $\mathcal{F}_1$ may be easily distinguished operationally by the fact that the corresponding physical devices look quite different, but there is no distinction between $\mathcal{F}_0$ and $\mathcal{F}_1$ from the point

[4] See page 25 of [19] for a nice discussion of this point

of view of the probabilistic predictions of the theory. We refer to the equivalence classes of events formed in this way as *transformations.* The following will mostly be concerned with transformations, rather than the underlying primitive events. Transformations with no input ports we will sometimes call *states*, and transformations with no output ports, *effects.* For system types A and B, the sets of transformations from A to B, states on A, and effects on B are denoted $\mathbf{Transf(A,B)}$, $\mathbf{St(A)}$, and $\mathbf{Eff(B)}$ respectively. Note that equivalence classes can be composed together in sequence and parallel in the obvious manner. A transformation T is said to be *reversible* if there exists a transformation T' in the theory such that $T \circ T' = I = T' \circ T$, where I is the identity transformation for the theory.

It is convenient to use the 'Dirac-like' notation $|\sigma_r)_A$, where r indexes outcomes of the classical pointer, to represent a state of system type A, and ${}_A(\lambda_r|$ to represent an effect on system type A. If the state $|\sigma_{r_1})_A$ is followed by the effect ${}_A(\lambda_{r_2}|$, the joint probability of obtaining outcome r_1 for the preparation and outcome r_2 for the measurement is given by

$$ {}_A(\lambda_{r_2}|\sigma_{r_1})_A := P(r_1, r_2). $$

In the following, we shall sometimes drop the input/output type label. A state $|\sigma_{r_1})_A$ can be identified with a function $\widehat{\sigma}_{r_1}$ from effects on A into probabilities as follows:

$$ \widehat{\sigma}_{r_1} : \quad {}_A(\lambda_{r_2}| \mapsto {}_A(\lambda_{r_2}|\sigma_{r_1})_A. $$

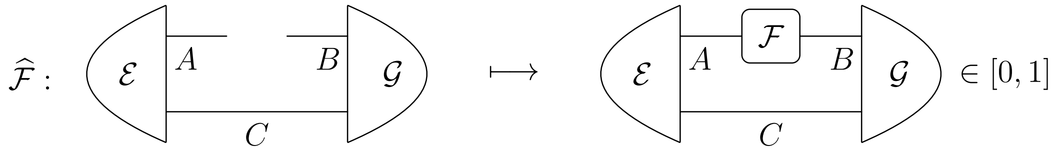
Since one can take real linear combinations of such functions, i.e. $\widehat{h} = \sum_i \alpha_i \widehat{\sigma}_r^i$, where $\alpha_i \in \mathbb{R}$, the set of states $\mathbf{St(A)}$ can be extended to (or be seen to generate) a real vector space, which we denote $\mathbf{V_A}$. It is clear by construction that the states span this vector space. In quantum theory, for example, states are positive operators, which span the real vector space $\mathbf{V_A}$ of Hermitian operators. Note that $\mathbf{St(A)}$ is a subset of $\mathbf{V_A}$, but is not in general a sub*space.* Similarly, an effect ${}_A(\lambda_{r_2}|$ can be identified with a function $\widehat{\lambda}_{r_2}$ from preparation events to probabilities:

$$ \widehat{\lambda}_{r_2} : \quad |\sigma_{r_1})_A \mapsto {}_A(\lambda_{r_2}|\sigma_{r_1})_A, $$

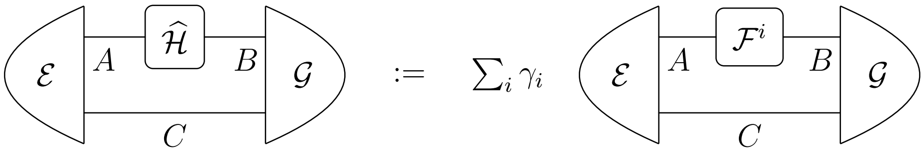
and the set of effects $\mathbf{Eff(A)}$ can be extended to a real vector space $\mathbf{V^A}$, where the vectors are again real linear combinations of the functions identified with the effects: $\widehat{g} = \sum_i \beta_i \widehat{\lambda}_r^i$. Again note that, by construction, the effects span this vector space. By construction, vectors in $\mathbf{V^A}$ act on vectors in $\mathbf{V_A}$ by linear extension as follows:

$$ \widehat{g}(\widehat{h}) := \sum_{ij} \alpha_i \beta_j \widehat{\lambda}_m^j \left(\widehat{\sigma_r}^i\right) = \sum_{ij} \alpha_i \beta_j (\lambda_m^j|\sigma_r^i). \tag{1.2} $$

A more general kind of transformation, from (possibly composite) system type A to (possibly composite) system type B, defines a function into probabilities, where the domain is now circuit fragments with the property that there are unconnected input and output ports, such that adding in a transformation of this type results in a closed circuit. Pictorially, a transformation $\mathcal{F}$ between systems A and B can be identified with a function $\widehat{\mathcal{F}}$ from appropriate circuits fragments into probabilities:



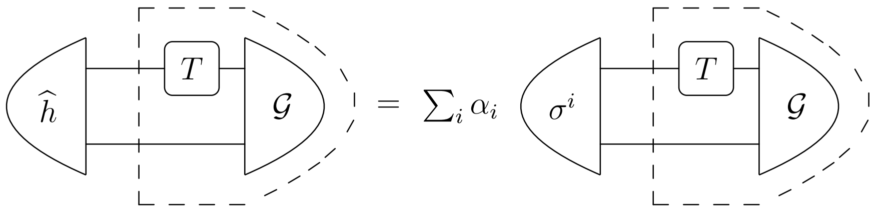
Again, this means that the set of transformations $\mathbf{Transf}(\mathbf{A},\mathbf{B})$ can be extended to a real vector space denoted $\mathbf{V}^{\mathbf{A}}_{\mathbf{B}}$, again spanned by the set of transformations, and that a function $\widehat{\mathcal{H}} = \sum \gamma_i \widehat{\mathcal{F}}^i$ acts on appropriate circuit fragments as:



Every transformation T from system A to system B induces a map $\widehat{T}_C$ from $\mathbf{V_{AC}}$ to $\mathbf{V_{BC}}$, for all systems C, defined by

$$|\sigma_r)_{AC} \in \mathbf{St(AC)} \mapsto (T \otimes I_C)|\sigma_r)_{AC} \in \mathbf{St(BC)},$$

where $(T \otimes I_C)|\sigma_r)_{AC}$ is the state of type BC, corresponding to composition of T with $|\sigma_r)_{AC}$, with I_C understood as an identity transformation (or the absence of any transformation) on system C. We will now show that the probabilistic structure, together with the definition of the vector spaces $\mathbf{V_{AC}}$ and $\mathbf{V_{BC}}$, implies that $\widehat{T}_C$ must be a linear map, for all systems C. Given some $\widehat{h} = \sum_i \alpha_i \widehat{\sigma}^i \in \mathbf{V_{AC}}$, the new function defined by $\widehat{T}_C \widehat{h}$ acts as:



where the equality follows from the conjunction of equation 1.2 with the fact that ${}_{BC}(\mathcal{G}|(T \otimes I_C)$—via compositionality—is an allowed effect, represented by the dashed box above. Moreover, this is true for any appropriately typed circuit fragment. Thus $\widehat{T}_C$ is a linear map:

$$\widehat{T}_C \widehat{h} = \widehat{T}_C \left(\sum_i^n \alpha_i \widehat{\sigma}_i \right) = \sum_i \alpha_i \left(T \otimes I_C \right) |\sigma_i)_{AC}.$$

Henceforth, we shall drop the distinction between a transformation $T \otimes I_C$ and the linear map corresponding to it, $\widehat{T}_C$. Throughout the thesis, we adopt the following assumption:

Finite tomography. *For each set* $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$*, there exist a finite and minimal set of appropriately typed circuit outcome fragments* $\{f_i\}_{i=1}^n$ *such that if*

$$f_i\left(T\right) = f_i \left(\sum_k \gamma_k \mathcal{F}^k \right) \qquad \forall i = 1, \dots, n,$$

for any $T \in \mathbf{Transf}(\mathbf{A}, \mathbf{B})$ *and any real linear combination of transformations from* $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$*,* $\sum_k \gamma_k \mathcal{F}^k \in \mathbf{V}_{\mathbf{B}}^{\mathbf{A}}$*, then* $T = \sum_k \gamma_k \mathcal{F}^k$ *as vectors in* $\mathbf{V}_{\mathbf{B}}^{\mathbf{A}}$*.*

In the above, minimality corresponds to the fact that no f_j is a linear combination of any of the remaining f_i for $\neq j$. Operationally, this corresponds to the fact that the possible measurement probabilities for f_j cannot be computed by knowing the measurement probabilities for the remaining f_i.

Operationally, finite tomography means that the state of every system (in general, each transformation between two systems) can be identified from the statistics of a finite number of finite-outcome measurements (in general, circuit fragments). The above statement is in fact slightly more general than this; it says that if one cannot distinguish a state (in general, transformation) from a specific linear combination of states (in general, transformations) with any fiducial measurement, then one cannot distinguish them with any measurement.

Finite tomography implies that, for every pair of system types A and B, and every transformation from A to B, $\mathbf{V}_{\mathbf{B}}^{\mathbf{A}}$ is finite dimensional, with dimension equal to the number fiducial measurements, as will be shown below. As a consequence, the vector space generated by effects on a system can be regarded as dual to the space of states, and vice versa: $\mathbf{V}^{\mathbf{A}} = (\mathbf{V}_{\mathbf{A}})^*$ and $\mathbf{V}_{\mathbf{A}} = (\mathbf{V}^{\mathbf{A}})^*$. In previous discussions of generalised probabilistic theories [15, 16], the finite dimensionality of these vector spaces was merely assumed from the outset. Our intent with finite tomography was to determine the operational content of this assertion. We will show, by providing an explicit example at the end of this section, that the requirement of distinguishing states from linear combinations of states in the statement of finite tomography is essential to deriving the finite dimensionality of the above vector spaces.

We now prove that finite tomography implies the finite dimensionality of vector spaces generated by transformations in the theory. Choose a set of n transformations $\{T^i\}_i$ with the property that they are linearly independent vectors in $\mathbf{V}_{\mathbf{B}}^{\mathbf{A}}$. That is, for any T^i, there does not exist any real coefficients λ_j such that $f_k(T^i) = f_k\left(\sum_{j\neq i}\lambda_j T^j\right) = \sum_{j\neq i}\lambda_j f_k(T^j)$ for all f_k. As every vector space (even infinite dimensional ones) has a basis, and transformations span the entire vector space by construction, linearly independent transformations must exist. Moreover, at least n linearly independent transformations must exists due to the minimality of the f_i's. As transformations span $\mathbf{V}_{\mathbf{B}}^{\mathbf{A}}$, we only need show that every transformation can be written as a real linear combination of the n T^i's. To characterise a transformation, we need only know its action on the fiducial set $\{f_i\}_{i=1}^n$. All we need to show is that, for a transformation $\mathcal{F}$, there exists a unique solution $\{\alpha_i\}$ to the following set of equations:

$$f_1(\mathcal{F}) = f_1\left(\sum_i^n \alpha_i T^i\right) = \sum_i^n \alpha_i f_1\left(T^i\right)$$

$$.$$

$$.$$

$$f_n(\mathcal{F}) = f_n\left(\sum_i^n \alpha_i T^i\right) = \sum_i^n \alpha_i f_n\left(T^i\right)$$

which can be rewritten as follows:

$$\begin{pmatrix} f_1(\mathcal{F}) \\ \vdots \\ f_n(\mathcal{F}) \end{pmatrix} = \begin{pmatrix} f_1(T^1) & \cdots & f_1(T^n) \\ \vdots & \vdots & \vdots \\ f_n(T^1) & \cdots & f_n(T^n) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

The above linear equations have a unique solution if the $n \times n$ matrix is invertible. This is equivalent to the columns of the matrix being linearly independent, i.e. to $\sum_i^n \lambda_i f_k(T^i) = 0$, for all $k = 1, \ldots, n$, implying $\lambda_i = 0$ for all i. By linearity we can rearrange this as follows $f_k(T^i) = -f_k\left(\sum_{j\neq i}\lambda_j T^j\right) = -\sum_{j\neq i}\lambda_j f_k(T^j)$. Finite tomography, in conjunction with a little rearranging, then implies $\sum_i^n \lambda_i T^i = 0$. As $\{T^i\}$ where chosen to be linearly independent, the required result follows.

An essential ingredient in the above proof was the fact that finite tomography guarantees the fiducial set is sufficient to distinguish[5] any two vectors from $\mathbf{V}_{\mathbf{B}}^{\mathbf{A}}$. In

[5] Indeed, given any two vectors, v, w, from $\mathbf{V}_{\mathbf{B}}^{\mathbf{A}}$, they can always be written as $v = \sum_i \alpha_i T^i$ and $w = \sum_i \beta_i T^i$ for transformations $\{T^i\}$ and real numbers $\{\alpha_i\}, \{\beta_i\}$. If one has $f_k(v) = f_k(w)$ for

other words, the fiducial set is *separating* for the vector space $\mathbf{V}_{\mathbf{B}}^{\mathbf{A}}$. It is interesting to wonder whether the full statement of finite tomography is needed for this, or if it is sufficient for the fiducial set to only separate $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$ in order to separate the vector space generated by $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$. If this were the case, we could derive the finite dimensionality of the vector space generated by $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$ from the requirement that the fiducial set distinguishes transformations, and nothing more.

We now provide an example showing that, in general, separating a spanning set is insufficient to separate the entire vector space it spans. Hence, the full statement of tomographic locality is required to ensure the finite dimensionality of the vector spaces considered above. Consider the following three linearly independent vectors which span $\mathbb{R}^3$:

$$\sigma_1 = \begin{pmatrix} 1 \\ 1 \\ 1/2 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The following set of two vectors from the dual space:

$$e_1 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$$

are sufficient to distinguish any of the above three vectors. For example e_2 distinguishes σ_1 from σ_2, as $e_2(\sigma_1) = e_2 \cdot \sigma_1 = 1/2 \neq 0 = e_2 \cdot \sigma_2 = e_2(\sigma_2)$, with $\cdot$ denoting the standard inner product on $\mathbb{R}^3$. Yet they are insufficient to distinguish σ_1 from the vector $\sigma_2 + \frac{1}{2}\sigma_3$.

In other works on generalised probabilistic theories, it is quite often assumed that the sets $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$, $\mathbf{St}(\mathbf{A})$, and $\mathbf{Eff}(\mathbf{B})$ are convex subsets of the corresponding vector spaces, the idea being that probabilistic mixtures of allowed transformations should also be allowed transformations. This work, however, doesn't need this assumption: the main constraints on sets of transformations, states and effects are closure under sequential and parallel composition. This (potential) lack of convexity is the reason fiducial measurements were required to distinguish not only states, but linear combinations of states.

Indeed, in the convex sets approach to generalised probabilistic theories [17, 18], fiducial circuit fragments are only required to distinguish allowed transformations in the theory. This is due to the fact that—in the convex sets approach—convex combinations of transformations are allowed transformations in the theory. Moroever,

all k from the fiducial set, one can use linearity of the f_k's and rearrange to arrive at $f_k(T^i) = f_k\left(\sum_i \frac{\beta_i}{\alpha_i} T^i - \sum_{j \neq i} \frac{\alpha_j}{\alpha_i} T^j\right)$. From this finite tomography implies $v = w$.

distinguishing convex combinations can be extended to distinguishing linear combinations (see appendix 1 of [18]). Hence, distinguishing transformations in the convex sets framework is equivalent to distinguishing transformations from linear combinations of transformations in the circuit framework.

1.3 Physical principles

We now introduce the physical principles which will play a role in subsequent chapters. A physical principle is any statement that can be made with only operational notions, such as preparations, outcome, experiment, etcetera.

1.3.1 Tomographic locality

Recall from section 1.2 that every transformation T_s from A to B induces a linear map from $\mathbf{V_A}$ to $\mathbf{V_B}$, defined by

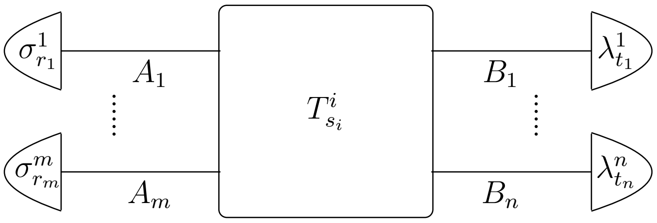
$$|\sigma_r)_A \in \mathbf{St(A)} \mapsto T_s|\sigma_r)_A \in \mathbf{St(B)}, \tag{1.3}$$

where $T_s|\sigma_r)_A$ is the state of type B, corresponding to composition of T_s with $|\sigma_r)_A$. Without further assumptions, however, this map is in general *not* sufficient to specify the transformation T_s. To see this, consider the situation in which the transformation T_s is applied to one half of a bipartite state $|\sigma)_{AC}$. The composition defines a bipartite state of type BC, which can be schematically represented $|\sigma')_{BC} = (T_s \otimes I_C)|\sigma)_{AC}$. The action of T_s on bipartite states of type AC induces a linear map from $\mathbf{V_{AC}}$ to $\mathbf{V_{BC}}$. In general, however, there need be no simple relationship between this map, and the above map from $\mathbf{V_A}$ to $\mathbf{V_B}$. Indeed, there need not be any simple relationship between the vector space $\mathbf{V_{AC}}$ and the vector spaces for the individual systems, $\mathbf{V_A}$ and $\mathbf{V_C}$. For each possible system type C, this structure is ultimately specified by the theory, via the assignments of probabilities to circuit outcomes[6].

The representation of transformations in a generalised probabilistic theory is greatly simplified by the assumption of *tomographic locality*. Roughly, a theory satisfies tomographic locality if every transformation can be fully characterized by local process tomography. More formally, consider transformations $T^1_{s_1}$ and $T^2_{s_2}$, both of

[6] The operational content of finite tomography is that there does at least exist a finite set of system types C, such that specification of the action of $T_s \otimes I_C$ on $\mathbf{V_{AC}}$ for each of the system types in this finite set is sufficient to characterise T_s.

which have input type $A_1 \cdots A_m$ and output type $B_1 \cdots B_n$. Consider circuit outcomes of the form



(1.4)

with corresponding probability $P^i(r_1 \ldots r_m, t_1 \ldots t_n, s_i)$, where $i \in \{1, 2\}$. Tomographic locality states that for all transformations $T^1_{s_1}$ and $T^2_{s_2}$ with matching input and output types, if

$$P^1(r_1 \ldots r_m, t_1 \ldots t_n, s_1) = P^2(r_1 \ldots r_m, t_1 \ldots t_n, s_2) \quad \forall |\sigma^1_{r_1}), \ldots, |\sigma^n_{r_m}), (\lambda^1_{t_1}|, \ldots, (\lambda^n_{t_n}|$$

then

$$T^1_{s_1} = T^2_{s_2}.$$

The circuit outcome fragment into which T_s was inserted in diagram 1.4—which corresponded to attaching a state to each input port and an effect to each output port—will be referred to as a *local circuit fragment*. The full statement of tomographic locality is slightly more general than what was discussed above; it says that, similar to the case of finite tomography, if one cannot distinguish a transformation from a specific linear combination of transformations by inserting them into any local circuit fragment, then one cannot distinguish them with any circuit fragment.

Principle 1. ***Tomographic locality:*** *If, for any transformation $T \in \mathbf{Transf}(\mathbf{A}, \mathbf{B})$ and any real linear combination of transformations $\sum_k \gamma_k \mathcal{F}^k$, one has*

$$l\,(T) = l\left(\sum_k \gamma_k \mathcal{F}^k\right) = \sum_k \gamma_k l\left(\mathcal{F}^k\right)$$

for all local circuit outcome fragments l, then $T = \sum_k \gamma_k \mathcal{F}^k$.

In the convex sets approach to generalised probabilistic theories [17, 18], the principle of tomographic locality is stated slightly differently. There, local circuit fragments are only required to distinguish allowed transformations in the theory. This is due to the fact that—in the convex sets approach—convex combinations of transformations are allowed transformations in the theory. Moroever, distinguishing convex combinations can be extended to distinguishing linear combinations (see appendix 1 of [18]). The (potential) lack of convexity in the circuit framework is the reason local circuit fragments are required to distinguish not only transformations, but linear combinations of transformations.

Violation of tomographic locality corresponds to the existence of global degrees of freedom that are not accessible to local measurements. A consequence of tomographic locality is that for a transformation with input type AB and output type CD, the corresponding real vector space has the form [17, 15, 16],

$$\mathbf{V}^{\mathbf{AB}}_{\mathbf{CD}} \cong \mathbf{V}^{\mathbf{A}} \otimes \mathbf{V}^{\mathbf{B}} \otimes \mathbf{V}_{\mathbf{C}} \otimes \mathbf{V}_{\mathbf{D}}, \tag{1.5}$$

where $\otimes$ here denotes the ordinary vector space tensor product (as opposed to the symbolic $\otimes$ used previously to denote parallel composition). In particular, for a bipartite state of type AC, the corresponding vector space $\mathbf{V}_{\mathbf{AC}} \cong \mathbf{V}_{\mathbf{A}} \otimes \mathbf{V}_{\mathbf{C}}$. This follows from the fact that local circuit outcome fragments are separating[7] for the vector space $\mathbf{V}^{\mathbf{AB}}_{\mathbf{CD}}$ and thus—due to finite tomography implying each vector space is finite dimensional—local circuit outcome fragments are spanning for the space dual to $\mathbf{V}^{\mathbf{AB}}_{\mathbf{CD}}$ (which in the case of states is just the vector space of effects) and thus the dual space decomposes as a tensor product. By finite dimensionality it follows that the original vector space also decomposes as a tensor product.

Furthermore, a transformation $T_s \in \mathbf{Transf}(\mathbf{A}, \mathbf{B})$ is completely specified by its action on $\mathbf{St}(\mathbf{A})$, hence T_s can be identified with the linear map defined by equation 1.3. When T_s acts on part of a bipartite state of type AC, the induced linear map $\mathbf{V}_{\mathbf{AC}} \to \mathbf{V}_{\mathbf{BC}}$ is given by $T_s \otimes I_C$, where again, the symbol $\otimes$ represents the ordinary vector space tensor product, and I_C is now the identity operator on the vector space $\mathbf{V}_{\mathbf{C}}$.

The formulation of tomographic locality in [15], referred to as local distinguishability by those authors, states that only transformations need be distinguished by local circuit fragments and does not ask for the ability to distinguish transformations from linear combinations of transformations. They then infer from this weaker notion of tomographic locality that, as local circuit fragments separate the set of transformations, they separate the vector space generated by that set. But, as we illustrated with an example at the end of section 1.2, separating a spanning set is in general insufficient to separate the vector space they span. Hence we have introduced a slightly more general version of tomographic locality here that ensures a vector space tensor product structure.

[7] Recall the difference between separating for the spanning set versus separating for vector space spanned by that set discussed at the end of section 1.2.

Fixing a basis for each system type, a transformation T with input AB and output CD can be written as a matrix

$$T = \sum_{i,j,k,l} M_{ij,kl}\big(\alpha_i^A \otimes \alpha_j^B \otimes \alpha_k^C \otimes \alpha_l^D\big),$$

where $M_{ij,kl} \in \mathbb{R}$, $\{\alpha_i^A\}$, $\{\alpha_j^B\}$ are bases for $\mathbf{V^A}$ and $\mathbf{V^B}$ respectively, and $\{\alpha_l^C\}$, $\{\alpha_m^D\}$ are bases for $\mathbf{V_C}$ and $\mathbf{V_D}$ respectively. The probability associated with a circuit outcome, e.g., of the form of diagram (1.1), can be written

$$M_{r_3}^3 \cdot (M_{r_2}^2 \otimes I_C) \cdot M_{r_1}^1,$$

where $M_{r_1}^1$ (a column vector) is the matrix form of the transformation corresponding to the event $\mathcal{E}_{r_1}$, $M_{r_2}^2$ corresponds to $\mathcal{F}_{r_2}$, $M_{r_3}^3$ (a row vector) corresponds to $\mathcal{G}_{r_3}$, and $\cdot$ is the standard inner product on finite dimensional real vector spaces.

1.3.2 Causality

A nice feature of the Pavia-Hardy framework we have described is that a basic assumption of causality is not implicit, but can be articulated explicitly and theories considered that do not satisfy this assumption. A generalised probabilistic theory is said to be *causal* if:

Principle 2. ***Causality:*** *The marginal probability of a preparation event is independent of the choice of which measurement follows the preparation.*

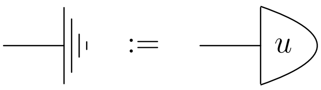
More formally, if $\{|\sigma_r)\}_{r\in X} \subset \mathbf{St(A)}$ are the states corresponding to a preparation test, consider the probability of outcome r, given that a subsequent measurement $\mathcal{E}$ corresponds to a set of effects $\{(\lambda_j|\}_{j\in Y}$:

$$P(r|\mathcal{E}) := \sum_{j\in Y} (\lambda_j|\sigma_r).$$

The theory is causal if for any system type A, any preparation test with outcome i, and any pair of measurements, $\mathcal{E}$ and $\mathcal{F}$, with input type A,

$$P(r|\mathcal{E}) = P(r|\mathcal{F}), \quad \forall r.$$

If circuits are thought of as having a temporal order, with tests later in the sequence occurring at a later time than tests earlier in the sequence, then the assumption of causality captures the intuitive notion of *no signalling from the future.* It was shown in [15] that a generalised probabilistic theory is causal if and only if for every system type A, there is a unique deterministic effect ${}_A(u|$. In this case, a measurement, with corresponding effects $\{(\lambda_j|\}_{j\in Y}$, satisfies $\sum_j (\lambda_j| = (u|$. A state $|\sigma)$ is normalised if and only if $(u|\sigma) = 1$. The unique deterministic effect provides a unique notion of *marginalisation* for multi-partite states. As the parallel composition of two single outcome tests is a single-outcome test, the effect ${}_A(u| \otimes_B (u|$ is deterministic for system AB. Hence, as causality implies the deterministic effect for any given system must be unique, we have ${}_A(u| \otimes_B (u| =_{AB} (u|$ for any systems A and B. Diagrammatically, we will denote the unique deterministic effect $(u|$ as follows:



The causality assumption also implies [15] a *no-signalling* principle for the states of the theory. That is, in a causal theory, if a test is performed on the A part of a composite system of type AB, then it is not possible to get information about which test was performed by only performing a test on the B part. (For an interesting extension of this idea to arbitrary causal networks, corresponding to circuits in the Pavia-Hardy framework, see [28].)

Although the idea of *no-signalling from the future* seems intuitive, there is nothing obviously pathological about generalised probabilistic theories that do not satisfy the causality assumption, as long as one does not try to define adaptive circuits, wherein a choice of later test can depend on an earlier outcome. Indeed, there is nothing about the framework as it stands that forces an interpretation of the circuits described previously as a sequence of tests applied in a temporal order which matches the order of tests in the circuit. Perhaps an entire closed circuit is set up in advance, and the pointers attain their final resting positions together, when a "go" button is pressed. An example of a non-causal theory will be presented in section 1.4 of this chapter.

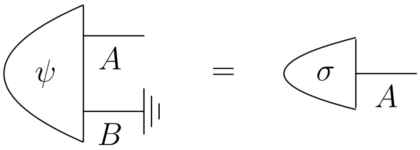
1.3.3 Purification and purity preservation

Before we can define purification and purity preservation, we need the notion of a *pure* transformation. Towards this end, consider the following. We say the laboratory device $\{\mathcal{U}_j\}_{j\in Y}$, where j indexes the positions of the classical pointer, is a *coarse-graining* of the device $\{\mathcal{E}_i\}_{i\in X}$ if there is a disjoint partition $\{X_j\}_{j\in Y}$ of X such that $\mathcal{U}_j = \sum_{i\in X_j} \mathcal{E}_i$. That is, coarse-graining arises when some outcomes of a laboratory device are joined together. The device $\{\mathcal{E}_i\}_{i\in X}$ is said to *refine* the device $\{\mathcal{U}_j\}_{j\in Y}$. A transformation T is said to be *pure* if and only if, for any test $\{\mathcal{U}_j\}_{j\in Y}$ containing T, and any test refining $\{\mathcal{U}_j\}_{j\in Y}$, $\{\mathcal{E}_i\}_{i\in X}$ say, one has $\mathcal{E}_i = p_i T$ for any $\mathcal{E}_i$ appearing in $T = \sum_{i \in X_j} \mathcal{E}_i$, where $\{p_i\}_i$ is a probability distribution. A pure[8] transformation is a process about which we have maximal information.

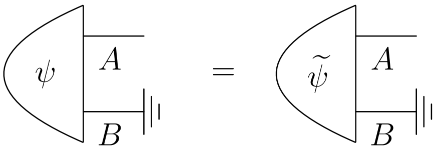
In particular, a state is pure if it does not arise as a *coarse-graining* of other states; a pure state is one for which we have as much information as possible. A state is *mixed* if it is not pure and it is *completely mixed* if any other state refines it. That is, $|c)$ is completely mixed if for any other state $|\rho)$, there exists a non-zero probability p such that $p|\rho)$ refines $|c)$.

We can now introduce the *purification* principle.

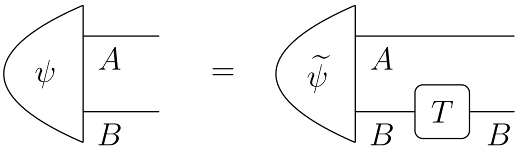
Principle 3. ***Purification:*** *Given a state $|\sigma)_A$ there exists a system B and a pure state $|\psi)_{AB}$ on AB such that $|\sigma)_A$ is the marginalisation of $|\psi)_{AB}$:*



Moreover, the purification $|\psi)_{AB}$ is unique up to reversible transformations on the purifying system, B. That is, if two states $|\psi)_{AB}$ and $|\widetilde{\psi})_{AB}$ purify $|\sigma)_A$



then there exists a reversible transformation T on system B such that



Purification can be thought of as saying that information can never be destroyed only discarded. Or alternatively, that—at some level—information is fundamentally conserved. Indeed, it states that any uncertainty or "mixedness" in a state is due to lack of knowledge of a suitable purifying system, or "environment". Purification is a strong principle and has many consequences. It implies that $\mathbf{St(B)}$ is *transitive*

[8] This definition of purity may be slightly misleading. Indeed, one might expect the identity transformation in classical probability theory to be pure, but this is not the case as $\mathbb{I} = \sum_i P_i$, where P_i are rank one projectors and $\{P_i\}_i$ forms a test. See [23, page 16] for a more in depth discussion on this point. Despite this, the notion of a pure state provided by this definition is indeed the correct notion one would expect.

for any system B. That is, given any two pure states from $\mathbf{St(B)}$, there exists a reversible transformation in the theory that maps one to the other. Indeed, every pure state on system B is a purification of any state on the trivial system, i.e. no system, and hence, by the uniqueness of purification up to reversible transformations on the purifying system, any two two pure states are connected by a reversible transformation. Purification—in conjunction with non-determinism[9]—also implies the existence of entangled, i.e. non-separable[10], states and a probabilistic protocol for teleportation and entanglement swapping [15]. Purification also implies the existence of a completely mixed state for any system [15], corresponding to the state left invariant by all reversible transformations on that system.

Principle 4. ***Purity Preservation:*** *Composition preserves purity. That is, sequential and parallel compositions of any pure transformations result in a pure transformation.*

Informally, purity preservation implies that if one has maximal knowledge about individual transformations, then one also have maximal knowledge about their composite. The above principles will play an important role in chapters 6 and 7.

1.3.4 Strong symmetry

States $\{|\sigma_i)\}_{i=1}^n$ are *perfectly distinguishable* if there exists a measurement, corresponding to effects $\{(e_i|\}_{i=1}^n$, such that $(e_i|\sigma_j) = \delta_{ij}$ for all i, j. Note that such an n-tuple of states can reliably encode an n-level classical system. Hence, sets of pure and perfectly distinguishable states can in some sense be thought of as perfect information carriers.

Principle 5. ***Strong symmetry:*** *A theory satisfies* strong symmetry *if for any two n-tuples of pure and perfectly distinguishable states $\{|\rho_1), \dots, |\rho_n)\}$, and $\{|\sigma_1), \dots, |\sigma_n)\}$, there exists a reversible transformation T such that $T|\rho_i) = |\sigma_i)$ for $i = 1, \dots, n$.*

Informally, strong symmetry can be thought of as saying that all information carriers of the same size are equivalent. Or, put differently, information is independent of the encoding media. The principle of strong symmetry implies the existence of non-trivial dynamics in any theory satisfying it. In chapter 4, we will mainly be concerned with two special cases of the above principle:

[9] i.e. the existence of measurement outcomes which occur with probability not equal to 0 or 1.

[10] A state is separable if it can be written as a convex combination of product states.

1. **Permutability**: A general theory satisfies *Permutability* if for any n-tuple of pure and perfectly distinguishable states and any permutation π of this n-tuple

$$\{|\rho_1), \ldots, |\rho_n)\} \quad \& \quad \{|\rho_{\pi(1)}), \ldots, |\rho_{\pi(n)})\},$$

there exists a reversible transformation T such that $T|\rho_i) = |\rho_{\pi(i)})$ for $i = 1, \ldots, n$.

2. **Bit-symmetry**: A theory satisfies *bit-symmetry* if for any two 2-tuples of pure and perfectly distinguishable states $\{|\rho_1), |\rho_2)\}, \{|\sigma_1), |\sigma_2)\}$, there exists a reversible transformation T such that $T|\rho_i) = |\sigma_i)$ for $i = 1, 2$.

Permutability is the special case of principle 5 where one of the sets of pure and perfectly distinguishable states is a permutation of the other. Bit-symmetry is the $n = 2$ case of principle 5.

1.4 Examples

The physical principles introduced in section 1.3 are, with the exception of causality and purification, all logically independent: generalised probabilistic theories satisfying any subset (including the empty subset) can be defined. We now provide examples of theories in this framework and discuss which principles are satisfied and violated in each case. Some interesting examples of generalised probabilistic theories that will not be discussed here are Spekkens toy model [123] and theories in which the set of states of a single system correspond to Euclidean hyperballs of dimension n [32, 100] (the $n = 3$ case of such theories corresponds to the Bloch ball of quantum theory).

1.4.1 Quantum theory

Finite-dimensional quantum theory provides a specific example of a theory that can be described in this framework. A system is associated with a complex Hilbert space, with the type of the system given by the dimension of the Hilbert space. States and effects are associated with positive operators, and transformations are associated with trace non-increasing completely positive maps. A test with no input ports corresponds to what is sometimes called a 'random source of quantum states', and is associated with positive operators $\{\rho_r\}$ such that $\sum_r \mathrm{Tr}(\rho_r) = 1$. When the test is performed, the probability that the classical pointer takes position r is given by $\mathrm{Tr}(\rho_r)$, and the quantum state that is prepared, conditioned on the pointer reading being r, is the normalised operator $\rho_r/\mathrm{Tr}(\rho_r)$. A test with no output ports is associated with a positive operator-valued measurement, that is a set of positive operators $\{E_i\}$ satisfying $\sum_i E_i = \mathbb{I}$. A test with both input and output ports is associated with a *quantum instrument*, that is a set of trace non-increasing completely positive maps, one for each value of the pointer reading r, that sum to a trace-preserving map. Given these associations, the standard rules of quantum theory allow the probability to be calculated for any circuit outcome.

Tomographic locality: For a system of type A, the vector space $\mathbf{V_A}$ is the real vector space of Hermitian operators, spanned by the density matrices. It is well known that quantum theory satisfies the assumption of tomographic locality. This follows from the way in which systems combine to form composite systems: a joint state is a positive operator acting on the tensor product of the Hilbert spaces associated with the individual systems. One can then check that the real vector spaces of Hermitian operators satisfy $\mathbf{V_{AB}} \cong \mathbf{V_A} \otimes \mathbf{V_B}$.

Causality: Quantum theory satisfies the causality assumption, as the probability of an event cannot depend on the choice of a measurement that is subsequently performed on the system. For a system associated with Hilbert space H, the unique deterministic effect, guaranteed to exist in a theory satisfying causality, is simply the identity operator $\mathbb{I}$ on H. Indeed, recall from the above discussion that $\sum_i E_i = \mathbb{I}$ for any positive operator-valued measurement $\{E_i\}$.

Purificiation: Quantum theory also satisfies the purification principle. Indeed, every mixed state on a finite dimensional system $\sum_i p_i |i\rangle\langle i|$ can be purified to a state $|\psi\rangle\langle\psi|$, where $|\psi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle$, by the introduction of a suitable extra system. Moreover, any other purification $|\widetilde{\psi}\rangle$ must satisfy $|\psi\rangle = (\mathbb{I} \otimes U)\,|\widetilde{\psi}\rangle$ with U a unitary transformation. Purification is standardly referred to by mathematicians as the Gelfand-Naimark-Segal construction [26, 27].

Purity preservation: In quantum theory, pure transformations correspond to completely positive maps with a single Kraus operator, that is a completely positive map $\mathcal{E}$ with $\mathcal{E}(\cdot) = E \cdot E^\dagger$. Clearly the sequential or parallel composition of pure quantum maps is also a completely positive map with only a single Kraus operator, i.e. $\mathcal{F}\mathcal{E}(\cdot) = (FE)\cdot(FE)^\dagger$ and $\mathcal{F} \otimes \mathcal{E}(\cdot) = (F \otimes E)\cdot(F \otimes E)^\dagger$. Hence quantum theory satisfies purity preservation.

Strong symmetry: Lastly, quantum theory also satisfies strong symmetry. The simplest example of this is the Hadamard transformation. Given the two sets of pure and perfectly distinguishable states, $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ and $\{|+\rangle\langle +|, |-\rangle\langle -|\}$, the

Hadamard transformation is a unitary transformation H which satisfies $H|0\rangle\langle 0|H^\dagger = |+\rangle\langle +|$ and $H|1\rangle\langle 1|H^\dagger = |-\rangle\langle -|$.

1.4.2 Classical probability theory

The classical theory of finite dimensional probability distributions and stochastic processes is also an example of a specific theory in this framework. A system is associated with a real vector space with the type corresponding to the dimension of said vector space, which can be thought of as the number of discrete outcomes of some test on that system. A system of size n contains n pure states corresponding to vectors

$$|s_i) = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

with a 1 in entry i and a zero everywhere else. Mixed states are convex combinations of pure states

$$|p) = \sum_i p_i|s_i) = \begin{pmatrix} p_1 \\ \vdots \\ p_i \\ \vdots \\ p_n \end{pmatrix}.$$

Pure effects $(e_j|$ are represented by vectors dual to the pure states, i.e. $(e_j|s_i) = \delta_{ij}$. This implies the pure states $\{|s_i)\}$ form a perfectly distinguishable set. Arbitrary effects correspond to sums—or coarse-graining's—of pure effects, under the constraint that the sum of all effects in a measurement must equal the unique deterministic effect $(u| = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}$. Transformations in the theory correspond to stochastic matrices, that is square matrices where the entries all lie in the interval $[0, 1]$ and the columns all sum to one. The pure transformations are compositions of rank one projectors and permutation matrices.

Classical theory can be thought of as a limiting case of quantum theory where each density operator is diagonal in the same basis. Classical theory can thus be seen to satisfy tomographic locality, causality, purity preservation, and strong symmetry[11]. Classical theory does not, however, satisfy the purification principle. Indeed, as was remarked in section 1.3, purification implies the existence of pure entangled states,

[11] As classical theory only has one set of pure and perfectly distinguishable states, $\{s_i\}$, strong symmetry is equivalent to permutability in this context.

which immediately rules out classical theory. In fact, there is only one manner in which a causal theory without entangled states can satisfy the purification principle: the theory must not contain mixed states. This necessarily implies the theory is deterministic, i.e. the probabilities of any measurement must be either 0 or 1.

An interesting consequence of the purification principle—in conjunction with causality and tomographic locality [15]—is that one can always model an irreversible process as a reversible evolution of the system along with an environment which can always be taken to be in a pure state. Interestingly, classical probability theory satisfies a similar dilation result: a stochastic process can be simulated in a certain manner by a reversible process. The price for such a simulation is that one needs an external source of randomness, which is provided by the environment. That is, in classical theory, one can't in general take the environment to be in a pure state. This constitutes one of the central differences between quantum and classical theory.

1.4.3 Real Hilbert space quantum theory

Quantum theory defined over real, rather than complex, Hilbert spaces supplies an example of a theory that does not satisfy tomographic locality. In real quantum theory, states and effects correspond to real symmetric matrices (which satisfy the same set of constraints as the quantum case discussed above), and transformations to completely positive maps that preserve the set of real symmetric matrices.

We will now provide an example of two distinct transformations between two 2-dimensional systems (sometimes referred to as *rebit's* [30]), which cannot be locally distinguished in the theory and so provides a violation of tomographic locality. Consider

$$T_1(\rho) = \frac{1}{2}\rho + \frac{1}{2}Y\rho Y, \quad \text{and} \quad T_2(\rho) = \frac{1}{2}\mathbb{I}\cdot \mathrm{Tr}(\rho),$$

where $Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ is the Pauli Y matrix and T_2 is a measure and prepare transformation that traces out the input state and prepares the maximally mixed state, $\frac{1}{2}\mathbb{I}$. Clearly the transformation $Y \cdot Y$ is allowed in the theory as :

$$Y\rho Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} a & b \\ b & 1-a \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \begin{pmatrix} 1-a & -b \\ -b & a \end{pmatrix}, \quad \forall a, b \in \mathbb{R}.$$

It is easy to see that $T_1(\rho) = T_2(\rho)$ for all real symmetric matrices ρ with trace one. Hence, one cannot distinguish these two transformations by inputting a single rebit and performing a single measurement on the output state. To distinguish T_1 from T_2, one has to evaluate them on one half of the Bell state $|\phi^+\rangle\langle\phi^+|$ and perform a joint measurement of the two output systems. That is, one has

$$T_1 \otimes I\left(|\phi^+\rangle\langle\phi^+|\right) = \frac{1}{2}|\phi^+\rangle\langle\phi^+| + \frac{1}{2}|\psi^-\rangle\langle\psi^-|, \quad \& \quad T_2 \otimes I\left(|\phi^+\rangle\langle\phi^+|\right) = \frac{1}{4}\mathbb{I}\otimes\mathbb{I}.$$

Performing the two-outcome measurement $\{|\phi^+\rangle\langle\phi^+|+|\psi^-\rangle\langle\psi^-|, |\phi^-\rangle\langle\phi^-|+|\psi^+\rangle\langle\psi^+|\}$, where $|\phi^\pm\rangle = (|00\rangle \pm |11\rangle)/\sqrt{2}$ and $|\psi^\pm\rangle = (|01\rangle \pm |10\rangle)/\sqrt{2}$, distinguishes these two states. Note that as

$$(T_1 \otimes I - T_2 \otimes I)\left(|\phi^+\rangle\langle\phi^+|\right) = \frac{1}{4}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{1.6}$$

we have $\mathrm{Tr}\left((T_1 \otimes I - T_2 \otimes I)(E \otimes F)\right) = 0$ for all real symmetric matrices E, F. Hence one cannot distinguish these two states with local measurements. Moreover, one can think of equation 1.6 as a global degree of freedom not acessable to local observers. It was shown by Hardy and Wooters in [31] that one only ever needs to perform joint measurements between at most two subsystems to distinguish any two states in real quantum theory.

1.4.4 PR boxes, Boxworld, and polygon theories

We now present an example of a theory distinct from any quantum or classical theory, which satisfies tomographic locality, causality, and purity preservation, but which violates the purification principle and strong symmetry. This theory is known as "Boxworld" [17, 7] and allows for arbitrarily strong nonlocal correlations, such as the PR box correlations of Popescu and Rohrlich [29] which maximally violate the CHSH inequality.

For the simplest non-trivial single system A in Boxworld, there are two choices of binary-outcome measurements, $\{_A(x_a|\}$ for $x, a \in \{0,1\}$. Here x is the bit denoting the two possible choices of measurement and a is the bit denoting the two possible outcomes of the chosen measurement, i.e the two measurements on system A are $\{_A(0_0|,_A(0_1|\}$ and $\{_A(1_0|,_A(1_1|\}$. There are four possible pure states that can be prepared on system A, which will be denoted by $|z,w)_A$, with $z, w \in \{0,1\}$. For measurement $\{_A(0_0|,_A(0_1|\}$ they satisfy $(0_b|z,w) = \delta_{bw}$ and for measurement $\{_A(1_0|,_A(1_1|\}$ they satisfy $(1_b|z,w) = \delta_{bz}$. States and measurements in this theory can produce correlations associated with the so-called Popescu-Rohrlich non-local box [29]. That is, for a bipartite system AB, there exist states $|\rho_{PR})_{AB}$ such that

$$(x_a|(y_b|\rho_{PR})_{AB} = \begin{cases} \frac{1}{2}, & \text{if } a \oplus b = xy, \\ 0, & \text{otherwise} \end{cases}$$

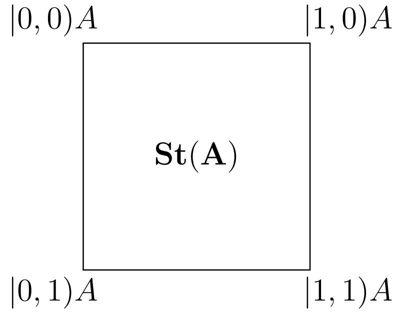


Fig. 1.2: State space of a single Boxworld system

where $\oplus$ represents addition modulo 2. The set of states $\mathbf{St}(\mathbf{A}) \subset \mathbb{R}^2$ contains all convex combinations of the four pure states $|x_a)_A$ and has a nice geometric manifestation as a square, where each vertex is associated with a pure state. This is depicted in Fig. 1.2 above. Reversible transformations for this single system correspond to the symmetry group of the above square.

It has been shown [7] that in Boxworld there is no teleportation, entanglement swapping, or dense coding. As discussed in section 1.3, the purification principle implies the existence of teleportation and entanglement swapping protocols. Hence Boxworld cannot satisfy purification. Moreover, Boxworld does not satisfy Bit-symmetry. Indeed, consider the two sets of pure and perfectly distinguishable states $\{|0,0)_A, |0,1)_A\}$—distinguished by measurement $\{_A(0_0|,_A(0_1|\}$—and $\{|0,0)_A, |1,1)_A\}$—distinguished by measurement $\{_A(1_0|,_A(1_1|\}$. As there is no symmetry of the square which takes $|0,1)_A$ to $|1,1)_A$ while leaving $|0,0)_A$ invariant, bit-symmetry cannot be satisfied. However, Boxworld does satisfy permutability as can be seen from the geometry of Fig. 1.2.

The particular manifestation of the set of states of a simple single system in Boxworld, depicted in Fig. 1.2, has motivated the introduction of theories in which the set of states of single systems correspond to regular polygons with n vertices [33, 34] (the case of $n = 4$ is the set of states in Boxworld depicted in Fig. 1.2). It was shown in [33] that polygon theories in which n is even violate the CHSH inequality more than T'sirelson's bound, but theories in which n is odd cannot violate T'sirelson's bound. Interestingly, as $n \to \infty$ the maximal violation of the CHSH inequality converges to the T'sirelson bound in both the even and odd cases. Theories with an even number of vertices do not satisfy the principle of strong symmetry, but theories with an odd number of vertices do satisfy strong symmetry [49].

1.4.5 A non-causal theory

An explicit theory that does not satisfy causality—first presented in [35]—will now be constructed and discussed. Systems are denoted schematically as $\mathrm{n} \triangleright \mathrm{m}$, where n, m are positive integers. Let Γ_n denote the set of all non-negative integers less than n, that is $\Gamma_n := \{0, \ldots, n-1\}$. The set of states of system $\mathrm{n} \triangleright \mathrm{m}$ is indexed by functions $f : X \to \Gamma_m$, where $X \subseteq \Gamma_n$. That is, $\mathbf{St}(\mathrm{n} \triangleright \mathrm{m}) := \{|\alpha_{f,X}) \mid f : X \to \Gamma_m \text{ and } X \subseteq \Gamma_n\}$. One can coarse-grain states corresponding to disjoint subsets $X, Y \subseteq \Gamma_n$ (i.e. $X \cap Y = \emptyset$) as follows: $|\alpha_{f,X}) + |\alpha_{g,Y}) := |\alpha_{h,Z})$, where $Z = X \cup Y$, $h|_X = f$ and $h|_Y = g$. Pure states thus correspond to $|\alpha_{f,\{i\}})$, where $f : \{i\} \to \Gamma_m$. We shall denote pure states as $|\alpha_{j,i})$, where $f(i) = j$. All states in $\mathbf{St}(\mathrm{n} \triangleright \mathrm{m})$ thus arise as coarse-graining's of pure states, and deterministic states correspond to states $|\alpha_{f,\Gamma_n})$.

We can now define the pure effects $(a_{k,l}|$ by $(a_{k,l}|\alpha_{j,i}) = \delta_{k,j}\delta_{l,i}$. One might expect that the entire set $\mathbf{Eff}(\mathrm{n} \triangleright \mathrm{m})$ can be built up by coarse-graining over the pure effects, as was the case for states. This turns out not to be the case. Indeed, consider the coarse-graining $(a_{k,l}| + (a_{k',l'}|$, with $k \neq k'$ and $l \neq l'$, and apply it to the state $|\alpha_{f,\{l,l'\}})$ where $f(l) = k$ and $f(l') = k'$. One has $\big((a_{k,l}| + (a_{k',l'}|\big)|\alpha_{f,\{l,l'\}}) = (a_{k,l}|\alpha_{k,l}) + (a_{k',l'}|\alpha_{k',l'}) = 2$, which clearly isn't allowed in any generalised probabilistic theory. Despite this, one can build up $\mathbf{Eff}(\mathrm{n} \triangleright \mathrm{m})$ via certain *allowed* coarse-gaining's of pure effects [35]. Indeed, it was shown[12] in [35] that $\mathbf{Eff}(\mathrm{n} \triangleright \mathrm{m}) := \{(a_{E,l}| \mid E \subseteq \Gamma_m \text{ and } l \in \Gamma_n\}$, hence every effect $(a_{E,l}|$ arises as a coarse-graining of pure effects $\sum_k (a_{k,l}|$, with $\cup_k\{k\} = E$. One can verify that $(a_{E,l}|\alpha_{f,X}) = \chi_X(l)\chi_E(f(l))$, where χ_S is the indicator function for the set S. Hence when applying the effect $(a_{E,l}|$ to the state $|\alpha_{f,X})$, one is essentially checking whether l is in the set X and $f(l)$ is in the set E.

Deterministic effects correspond to the elements $(e_l| := (a_{\Gamma_m,l}|$, note that there are n of them—one for each $l \in \Gamma_n$. Recall that the causality principle implies each system has a unique deterministic effect. Hence, this theory violates causality. We now present an explicit situation in which the 'acausality' of this theory is manifest. Consider the system $2 \triangleright 2$ and the preparation test $\mathcal{P} = \{|\alpha_{0,0}), |\alpha_{1,1})\}$, which we could have written as $\{|\alpha_{\mathrm{Id},\{i\}})\}_{i=0,1}$ with Id the identity function. Consider the two measurements $\mathcal{E} = \{(a_{0,0}|, (a_{1,0}|\}$ and $\mathcal{F} = \{(a_{0,1}|, (a_{1,1}|\}$. The ability to prepare the state $|\alpha_{0,0})$ will now be seen to depend on the choice of which measurement "follows"

[12] The authors of [35] were only interested in constructing a *deterministic* theory, i.e. a theory in which the probabilities of measurement outcomes are either 0 or 1. One could however take the convex closure of the sets of states and effects if one wanted to introduce some randomness.

$\mathcal{P}$: $\mathcal{E}$ or $\mathcal{F}$. Indeed,

$$\Pr(0,0|\mathcal{E}) = \sum_{k=0}^{1}(a_{k,0}|\alpha_{0,0}) = 1 \neq 0 = \sum_{k=0}^{1}(a_{k,1}|\alpha_{0,0}) = \Pr(0,0|\mathcal{F}).$$

Hence the constructed theory violates causality.

Composite systems are defined as $(\mathrm{n} \triangleright \mathrm{m})\,(\mathrm{n}' \triangleright \mathrm{m}') := \mathrm{x} \triangleright \mathrm{y}$, with $x = n \cdot n'$ and $y = m{\cdot}m'$. The pure states from $\mathbf{St}(\mathrm{x} \triangleright \mathrm{y})$ are built by taking the parallel composition of the pure states from $\mathbf{St}(\mathrm{n} \triangleright \mathrm{m})$ and $\mathbf{St}(\mathrm{n}' \triangleright \mathrm{m}')$, where now $\Gamma_x = \Gamma_n \times \Gamma_{n'}$ and $\Gamma_y = \Gamma_m \times \Gamma_{m'}$. The full set of states is then built by considering coarse-graining's of the pure states. It was shown in [35] that these composites satisfy tomographic locality. Lastly, the lack of a unique deterministic effect implies there is no unique way to marginalise over composite states. This fact directly leads to the ability to transmit signals instantaneously, see [35] for the details.

Chapter 2

Computation in generalised probabilistic theories

As discussed in the introduction, one of the major conceptual breakthroughs in physics over the past thirty years has been the realisation that quantum theory appears to offer dramatic advantages [1] for various information-processing tasks and computation in particular [64, 63, 1, 80, 81, 82, 83, 84, 38, 134, 149, 150]. This raises the general question of what relationships exist between physical principles, which a theory like quantum theory may or may not satisfy, and information theoretic advantages. As was discussed at the start of chapter 1, much progress has already been made in understanding the connections between physical principles and some tasks, such as cryptography and communication complexity problems. By comparison, relatively little has been learned about the connections between physical principles and computation.

It was shown in [11, 95, 96] that Boxworld has no non-trivial reversible dynamics and, hence, any reversible computation in Boxworld can be efficiently simulated on a classical computer. Aside from this result, most previous investigations into computation beyond the usual quantum formalism have focused on non-standard theories involving modifications of quantum theory. These theories often appear to have immense computational power and entail unreasonable physical consequences. For example, non-linear quantum theory appears to be able to solve[1] **NP**-complete problems in polynomial time [12], as does quantum theory in the presence of closed timelike curves [13, 59]. Aaronson has considered other modifications of quantum theory, such as a hidden variable model in which the history of hidden states can be read out by the observer [14], and—together with collaborators in [146]—a model in

[1] See appendix A for a rigorous definiton of the class **NP**

which one is given the ability to perform certain (unphysical) non-collapsing measurements. Both of these models have been shown to entail computational speed-ups over the usual quantum formalism. Lastly, Bao et al. [147] have investigated computation in modifications of quantum theory suggested by the black hole information loss paradox and have shown the ability to signal faster than light in such theories is intimately linked to a speed-up over standard quantum theory in searching an unstructured database.

In this chapter, we begin our investigation into computation in the framework of generalised probabilistic theories, which we introduced in chapter 1. As discussed previously, theories within this framework can be described that are different from classical or quantum theories, but which nonetheless make good operational sense and do not involve peculiarities like closed timelike curves. This framework suggests a natural model of computation, analogous to the classical and quantum circuit models, which we shall describe in section 2.1.

The strongest known non-relativised upper bound for the power of quantum computation—first proved by Fortnow and Rogers in [45]—is that the class **BQP** of problems efficiently solvable by a quantum computer is contained in the classical complexity class **AWPP**. This means that a quantum computer cannot solve any problem outside of **AWPP**, but it is unknown whether it can solve every problem contained in this class. The class **AWPP** has a slightly obscure definition, but is well known to be contained in **PP**, hence **PSPACE**. See appendix A for the definitions of all complexity classes mentioned in this thesis. Section 2.2 shows that the same result holds for any theory that satisfies the principle of tomographic locality, introduced in section 1.3 of chapter 1. That is, if the complexity class of problems that can be efficiently solved by a specific theory **G** is denoted schematically **BGP**, then for tomographically local theories, $\mathbf{BGP} \subseteq \mathbf{AWPP}$. Once suitable definitions are in place, the proof is essentially the same as the proof for the quantum case: the idea is that this proof can be cast in a theory-independent manner, and be seen to follow from a very minimal set of assumptions on the structure of a physical theory. In fact, the containment $\mathbf{BGP} \subseteq \mathbf{AWPP}$ still holds even in the absence of the principle of causality, also introduced in chapter 1.

It was suggested in [17] that, in some sense, quantum theory achieves an optimal balance between its set of states and its dynamics, and that this balance implies that quantum theory is powerful for computation by comparison with most theories in the space of operational theories. Although the status of this suggestion is unknown, it turns out to be exactly correct in the context of a world allowing post-selection of measurement outcomes. Aaronson showed that the class of problems efficiently solvable by a quantum computer with the ability to post-select measurement outcomes is equal to the class **PP** [65]. Section 2.3 extends the idea of computation with post-selection to general theories, and shows that given tomographic locality, problems efficiently solvable by any theory with post-selection are contained in **PP**. In other words: any problem efficiently solvable in a tomographically local theory with post-selection, is also efficiently solvable by a quantum computer with post-selection.

Finally, oracles play a special role in quantum computation, forming the basis of most known computational speed-ups over classical computation. Section 2.4 discusses the problem of defining a sensible notion of oracle in the general framework, which reduces to the standard definition in quantum theory. This problem may not have a solution that is completely general, hence we introduce here instead a notion of "classical oracle" that can be defined in any theory that satisfies the causality principle. There then exists a classical oracle such that relative to this oracle, **NP** is not contained in **BGP** for any theory **G** satisfying tomographic locality and causality. However, we do show in chapter 6 that such oracles exist in any theory satisfying sufficient physical principles.

The proofs of all theorems in this chapter will be presented in section 2.5.

2.1 The computational model

2.1.1 Uniform circuits

In chapter 1 we saw that in a generalised probabilistic theory, one can draw circuits representing the connections of physical devices in an experiment, and the specific events that took place in said experiment. These circuits provide a natural model of computation, based on the classical and quantum circuit models. However, a good notion of *efficient* computation needs a definition of a *uniform family of circuits* in a generalised probabilistic theory.

In the standard, classical or quantum, circuit model, a circuit family $\{C_n\} = \{C_1, C_2, \dots\}$ consists of a sequence of circuits, each indexed by a positive integer n, denoting the input system size, where C_n is the circuit corresponding to a problem instance of size n. In a poly-size circuit family, the number of gates in C_n is bounded by a polynomial in n, and the circuit family is uniform if a Turing machine can output a description of C_n in time bounded by a polynomial in n.

In a generalised probabilistic theory, there is no reason to assume that a circuit must have the form of a number of gates[2] acting on some input, where the input preparation encodes the problem instance—recall that in this chapter we are not necessarily assuming the principle of causality, in which case a circuit does not have a preferred direction. Instead, we allow the entire circuit to encode the problem instance, defining a circuit family as a set $\{C_x\}$ such that each circuit is indexed by a classical string $x = x_1 x_2 \dots x_n$. A circuit family is poly-size if the number of gates is bounded by a polynomial in $|x|$. For a particular generalised probabilistic theory it might not be the case that bipartite and single system transformations together are universal for computation, as they are in classical and quantum computation. Hence for any k, l, a circuit might involve gates with k input systems and l output systems. In general, it might be the case that no finite gate set is universal for computation. Nonetheless, we will impose as a requirement of uniformity that any uniform circuit family is associated with a finite gate set[3], such that each circuit in the family is built from elements of that set. It follows that the number of distinct system types appearing in a uniform circuit family is also finite.

A further requirement for a circuit family to be uniform takes the form of a constraint on the entries of the matrices representing the transformations that appear in the finite gate set—otherwise, it may be possible to smuggle hard to compute quantities into the computation. There must exist some fixed choice of basis of $\mathbf{V_A}$ for each system A, such that a Turing machine can efficiently compute approximations to the entries of the matrices relative to these bases. We require that for any matrix entry $(M)_{ij}$, and any ϵ, a Turing machine can output a rational number, within ϵ of $(M)_{ij}$, in time bounded by a polynomial in $\log(\frac{1}{\epsilon})$. We may motivate this by recalling that gates correspond to operational devices; an experimenter with access to devices governed by some generalised probabilistic theory may only be able to characterise them tomographically to finite precision—and the features of a probabilistic theory should not be sensitive to precision issues which are inaccessible to experiment. Indeed, it makes sense to assume that an experimenter with access to devices governed by some theory cannot align, or employ, them with arbitrary accuracy.

Finally, for a circuit family $\{C_x\}$ to be uniform, there must be a Turing machine that, acting on input x, outputs a classical description of C_x in time bounded by a polynomial in $|x|$.

[2] When discussing computation, the terms 'device' and 'gate' will be used interchangeably

[3] One might instead consider a uniformity condition where the number of permitted gates grows with circuit size, as in [58, §3.3 A]; however, we do not consider such a condition here.

The notion of a poly-size uniform circuit family $\{C_x\}$ can be summarised as follows:

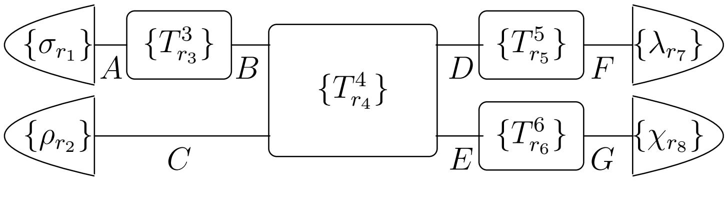
1. The number of gates in the circuit C_x is bounded by a polynomial in $|x|$.
2. There is a finite gate set $\mathcal{G}$, such that each circuit in the family is built from elements of $\mathcal{G}$.
3. For each type of system, there is a fixed choice of basis, relative to which transformations are associated with matrices. Given the matrix M representing (a particular outcome of) a gate in $\mathcal{G}$, a Turing machine can output a matrix $\widetilde{M}$ with rational entries, such that $|(M - \widetilde{M})_{ij}| \leq \epsilon$, in time polynomial in $\log(1/\epsilon)$.
4. There is a Turing machine that, acting on input $x = x_1 x_2 \ldots x_n$, outputs a classical description of C_x in time bounded by a polynomial in $|x|$.

Note that the notion of uniformity presented in this section is also required to define and discuss efficient classical and quantum computation, see [60] for an in-depth discussion of this point.

2.1.2 Acceptance criterion

Now that we have defined a uniform family of circuits, we need to discuss the acceptance criterion. In quantum computation it is known that performing intermediate measurements during the computation does not increase the computational power. So, without loss of generality, all measurements can be postponed until the end of the computation. A quantum computer can be defined to accept an input string x if the outcome of a computational basis measurement on the first outcome qubit is $|0\rangle$. In a general theory however, it need not be the case that all measurements can be postponed until the end of the computation without loss of generality, hence the acceptance criterion should reflect this.

The way in which a generalised probabilistic theory solves a problem might be imagined as follows. First, given the input string x, the circuit C_x is designed and built by composing gates from the fixed finite gate set sequentially and in parallel according to the description. An example of such a circuit is depicted schematically below.



Once the circuit is built, the computation can be run. At the end of a run, each gate has a classical outcome associated with it, where the theory defines a joint probability for these outcomes. For the example above, the joint probability is given by

$$P(r_1, \ldots, r_8) = (\chi_{r_8}|(\lambda_{r_7}|\big(T^6_{r_6} \otimes T^5_{r_5}\big)T^4_{r_4}\big(T^3_{r_3} \otimes I_C\big)|\rho_{r_2})|\sigma_{r_1}).$$

Denoting the string of observed outcomes by $z = r_1 \ldots r_8$, the final output of the computation will be given by a function of the observed outcomes $a(z) \in \{0, 1\}$, where there must exist a Turing machine that computes a in time polynomial in the length of the input $|x|$. The probability that a computation accepts the input string x is therefore given by

$$P_x(\text{accept}) = \sum_{z|a(z)=0} P(z),$$

where the sum ranges over all possible outcome strings of the circuit C_x.

2.1.3 Efficient computation

The class of problems that can be solved efficiently in a generalised probabilistic theory can be defined as follows.

Definition 2.1. *For a generalised probabilistic theory* **G**, *a language* $\mathcal{L}$ *is in the class* **BGP** *if there exists a poly-sized uniform family of circuits in* **G**, *and an efficient acceptance criterion, such that*

1. $x \in \mathcal{L}$ *is accepted with probability at least* $\frac{2}{3}$.

2. $x \notin \mathcal{L}$ *is accepted with probability at most* $\frac{1}{3}$.

As ever, the choice of the constant 2/3 is arbitrary. Any fixed constant k, $1/2 < k < 1$ would serve equally well[4]. Indeed, note that each uniform circuit (with an efficient acceptance condition) defines a random variable that maps circuit outcomes to the set {**accept, reject**} and so one can regard multiple repetitions of a computation as a collection of i.i.d. random variables (independence follows from the definition of the probabilistic structure given in chapter 1; specifically that the sequential or parallel composition of two events corresponding to outcomes of closed circuits define independent probability distributions). This fact is independent of the form of a particular theory and so holds true for all theories in the framework. Taking

[4] As indeed would any constant bounded away from 1/2 by an inverse polynomial in the size of the input.

this fact in conjunction with the definition of **BGP** and applying the Chernoff bound provides the required result. See [1, p.154] for more discussion of the quantum case.

For a specified **G**, the class **BGP** is the natural analogue of **BPP** for probabilistic classical computation, and **BQP** for quantum computation. Indeed, **BGP** reduces to **BPP** or **BQP** in the case that the theory **G** is in fact the classical or quantum theory. See, e.g., [36] for a proof that quantum circuits with mixed states and completely positive maps are equivalent in computational power to standard quantum circuits with pure states and unitary transformations.

Note that the way in which the acceptance criterion is defined implies that $\mathbf{P} \subseteq \mathbf{BGP}$, for (almost) every generalised probabilistic theory **G**. This is a consequence of the fact that the final output is a function $a(z)$ of the string of observed events z, and the only constraint on a is that it can be efficiently computed by a Turing machine. Degenerate cases provide exceptions to this—consider, e.g., any theory such that all transformations are deterministic, i.e., the outcome set of any circuit is the singleton set. One could remove these degenerate cases by generalising the acceptance function $a(.)$ so that it depend on both the outcome string z and the input string x. Of course, the fact that $\mathbf{P} \subseteq \mathbf{BGP}$ does not have much to do with the intrinsic computational power of a generalised probabilistic theory, but is an artefact of the acceptance criterion—it might be interesting to weaken this criterion so that computation in theories intrinsically weaker than classical can be explored.

2.2 Upper bounds on computational power

Using the above definitions of uniform circuit families, and acceptance of an input, the following upper bound on the computational power of any generalised probabilistic theory can be obtained. The main assumption—in addition to those involved in uniformity—is that tomographic locality holds. Note that the result does not require the causality assumption.

Theorem 2.2. *For any generalised probabilistic theory* **G** *satisfying tomographic locality*

$$\mathbf{BGP} \subseteq \mathbf{AWPP} \subseteq \mathbf{PP} \subseteq \mathbf{PSPACE}.$$

Here, **PSPACE** consists of those problems that, roughly speaking, can be solved by a classical computer using a polynomial amount of memory. **PP** stands for Probabilistic Polynomial time, which roughly speaking, contains those problems that can be solved by a probabilistic classical computer that must get the answer right with probability $> 1/2$. The probability does not need to be bounded away from $1/2$, indeed may be greater than $1/2$ only by an exponentially small amount, hence **PP** contains problems that are not thought to be efficiently solvable by a classical random computer. See appendix A for the formal definitions of all complexity classes discussed in this thesis. **AWPP** stands for Almost-Wide Probabilistic Polynomial time, and it is known that **AWPP** $\subseteq$ **PP** [45]. The best known upper bound for the class of efficient quantum computations, shown by Fortnow and Rogers in [45], similarly states that **BQP** $\subseteq$ **AWPP**.

To define the class **AWPP**, the notion of a **GapP** function must be introduced. Given a polynomial-time non-deterministic Turing machine M and input string x, denote by $M_{acc}(x)$ the number of accepting computation paths of M given input x, and by $M_{rej}(x)$ the number of rejecting computation paths of M given x. A function $f : \{0,1\}^* \to \mathbb{Z}$ is a **GapP** function if there exists a polynomial-time non-deterministic Turing machine M such that $f(x) = M_{acc}(x) - M_{rej}(x)$ for all input strings x. The class **AWPP** can be defined as follows [53].

Definition 2.3. *The class* **AWPP** *consists of those languages* $\mathcal{L}$ *such that there exists a* **GapP** *function* f*, and a polynomial* r *such that*

1. *If* $x \in \mathcal{L}$ *then* $2/3 \leq f(x)/2^{r(|x|)} \leq 1$,
2. *if* $x \notin \mathcal{L}$ *then* $0 \leq f(x)/2^{r(|x|)} \leq 1/3$.

Once the appropriate definitions for generalised probabilistic theories are in place, the proof of Theorem 2.2 is a fairly straightforward extension of similar proofs for the quantum case, and is presented in subsection 2.5.2.

Although formal proofs are relegated to section 2.5, it is useful to sketch the proof that **BGP** $\subseteq$ **PSPACE** in order to provide intuition about how the physical principles underlying generalised probabilistic theories lead to computational bounds.

Sketch proof. Consider a general circuit C_T, with $q(|T|)$ gates. Tensoring these gates with identity transformations on systems on which they do not act, and padding them with rows and columns of zeros, results in a sequence of square matrices $M^{r_q,q}, \ldots, M^{r_1,1}$, where $M^{r_n,n}$ is the matrix representing the r_n^{th} outcome of the n^{th} gate. This can be done in such a way that the probability for outcome $z = r_1 \ldots r_q$, is given by

$$b^T M^{r_q,q} \cdots M^{r_2,2} M^{r_1,1} b = \sum_{\{i_1,\ldots,i_{q-1}\}} M^{r_q,q}_{1 i_{q-1}} \cdots M^{r_2,2}_{i_2 i_1} M^{r_1,1}_{i_1 1}$$

where b is the vector $b = (1, 0, \ldots, 0)$ and b^T is its transpose. The output probability is a sum of exponentially many terms, but each term is a product of polynomially many numbers, each of which can be efficiently calculated. So a classical Turing machine can calculate each term in the sum, one after the next, keeping a running total. This requires only polynomial-sized memory. □

This proof relies on the ability to decompose the acceptance probability of the computation in a form reminiscent of a (discrete) Feynman path integral. This is a consequence of the fact that transformations in a generalised probabilistic theory are linear, and thus have a matrix representation. It is pertinent then to recall where this linearity comes from. When we introduced generalised probabilistic theories in chapter 1, we associated states (respectively, effects) with functions taking effects (respectively, states) to probabilities. As one can take linear combinations of such functions, this induces a linear structure on the set of states (respectively, effects). Thus the linear structure of generalised probabilistic theories arises from the requirement that a physical theory should be able to give probabilistic predictions about the occurrence of possible outcomes.

Aside from linearity, a further requirement of the proof is the ability to compute efficiently the entries in the matrices representing the transformations applied in parallel in a specific circuit. Recall from section 1.3.1 in chapter 1 that, in a theory satisfying tomographic locality, a transformation $\mathcal{E} \in \textbf{Transf}(\mathbf{A}, \mathbf{B})$ is completely specified by its action on $\mathbf{St}(\mathbf{A})$, and so the matrix representing transformations applied in parallel can be easily calculated by taking the tensor product of the matrices representing each individual transformation. This is not the case in a theory without tomographic locality, where the tensor product structure may not hold. If a transformation from $\mathbf{A}$ to $\mathbf{B}$ acts on one half of a system $\mathbf{AC}$, there may be no simple way to relate the linear map $\mathbf{St}(\mathbf{AC}) \to \mathbf{St}(\mathbf{BC})$ to the action of the transformation when it is applied to a system $\mathbf{A}$ on its own, or indeed to a joint system $\mathbf{AC}'$. There may therefore be no efficient way of computing matrix elements corresponding to a transformation considered as part of a circuit of arbitrary size.

Recall from chapter 1 that one can think of a violation of tomographic locality as corresponding to the existence of “global degrees of freedom” not accessible to local measurements. As the uniformity condition only imposes constraints on the matrices associated to gates from the finite gate set $\mathcal{G}$, if there doesn’t exist an efficient procedure for computing the matrices associated to the parallel compositions of gates from $\mathcal{G}$ one could in principle encode answers to computationally hard problems in these global degrees of freedom. While tomographic locality implies such an efficient procedure it may not be the simplest principle which does so. An interesting direction for future work might be to weaken the assumption of tomographic locality such that the results of theorem 2.2 still go through. Real Hilbert space quantum theory, discussed at the end of chapter 1, provides an example of a theory without tomographic locality for which the above bounds hold, since there is an efficient way of calculating relevant matrix entries.

2.3 Post-selection and generalised probabilisitic theories

In [65] Aaronson introduced the notion of *post-selected* quantum circuits. These are quantum circuits which, in addition to having a specified qubit, on which a computational basis measurement will be made to provide the outcome, have an additional qubit on which a measurement can be performed such that we can post-select on the outcome. Instead of sampling the measurement result r directly from the computational outcome qubit according to the distribution $P(r)$, only those runs of the computation are counted for which a measurement on the post-selected qubit yields the outcome $s = 0$. The outcome distribution for the computation is taken to be the conditional distribution $P(r|s = 0)$. An extra technical condition is needed, which is that there exists a constant D and polynomial w such that $P(S = 0) \geq 1/D^{w(|x|)}$, i.e., we can only post-select on at most exponentially-unlikely outcomes[5].

Definition 2.4. *A language $\mathcal{L}$ is in the class* **PostBQP** *if there is a polynomially-sized uniform quantum circuit family, where each circuit has a computational outcome qubit and a post-selected qubit, such that when computational basis measurements are performed on these qubits, with respective outcomes r and s,*

- *There exists a constant D and polynomial w such that $P(s = 0) \geq 1/D^{w(|x|)}$*
- *If $x \in \mathcal{L}$ then $P(r = 0|s = 0) \geq \frac{2}{3}$*
- *If $x \notin \mathcal{L}$ then $P(r = 0|s = 0) \leq \frac{1}{3}$*

Aaronson showed in [65] that **PostBQP** = **PP**. Combining this with Theorem 2.2 gives

[5] This extra condition was missing from Aaronson's original paper on **PostBQP**, but is needed for the definition of **PostBQP** to be independent of a choice of quantum gate set; see Section 2.5 of [37]. We thank Scott Aaronson for some very interesting discussions concerning this point.

Theorem 2.5. *For any generalised probabilistic theory* $\mathbf{G}$*,*

$$\mathbf{BGP} \subseteq \mathbf{PostBQP}.$$

Roughly speaking, a post-selecting quantum computer can simulate computation in any other theory satisfying tomographic locality. One can also define a notion of generalised circuits with post-selection on at most exponentially-unlikely outcomes. These are poly-sized uniform circuits in a generalised probabilistic theory, where the probability of acceptance is conditioned on the circuit outcome z lying in a (polytime computable) subset of all possible values of z. More formally:

Definition 2.6. *A language* $\mathcal{L}$ *is in the class* **PostBGP** *if there is a poly-sized uniform circuit family in that theory and an efficient acceptance condition, such that*

1. *There exists a constant* D *and polynomial* w *such that* $P(z \in S) \geq 1/D^{w(|x|)}$
2. *If* $x \in \mathcal{L}$ *then* $P_x(\text{accept}|z \in S) \geq \frac{2}{3}$
3. *If* $x \notin \mathcal{L}$ *then* $P_x(\text{accept}|z \in S) \leq \frac{1}{3}$

where z *is the circuit outcome,* S *is a subset of all possible circuit outcomes and* $z \in S$ *can be checked by a Turing machine in time polynomial in* $|x|$*.*

Given the above definition, one can now state the following theorem.

Theorem 2.7. *For any generalised probabilistic theory* $\mathbf{G}$*,*

$$\mathbf{PostBGP} \subseteq \mathbf{PP}.$$

The proof is in section 2.7. Combining this with Aaronson's result yields:

Corollary 2.8. *For any generalised probabilistic theory* $\mathbf{G}$*,*

$$\mathbf{PostBGP} \subseteq \mathbf{PostBQP}.$$

So, in a world in which we can post-select on at most exponentially-unlikely events, quantum theory is optimal for computation in the space of all tomographically local theories. Note that the class of problems efficiently solvable on a probabilistic classical computer with the power of post-selection is unlikely to be as large as **PP**: it was shown in [38] that if this class, denoted $\mathbf{BPP_{path}}$, is equal to **PP**, then the polynomial hierarchy collapses to the third level.

It was suggested in [17] (see also [49]) that quantum theory in some sense achieves an optimal balance between the sets of available states and dynamics, in such a way that quantum theory is optimal, or at least powerful, for computation relative to the class of generalised probabilistic theories. It is interesting to ask whether Corollary 2.8 can be seen as *evidence* in favour of this idea. The following considerations show that caution is needed. Consider, for example, the class **IQP** [38], of restricted quantum computations where the only gates allowed in a circuit are diagonal in the $\{|+\rangle, |-\rangle\}$ basis. Clearly $\mathbf{IQP} \subseteq \mathbf{BQP}$, but it is unlikely that $\mathbf{BQP} \subseteq \mathbf{IQP}$. However, it was shown in [38] that $\mathbf{PostIQP} = \mathbf{PP} = \mathbf{PostBQP}$. So, while $\mathbf{PostBQP} \subseteq \mathbf{PostIQP}$, it is not believed to be the case that $\mathbf{BQP} \subseteq \mathbf{IQP}$. Alternatively, consider the class of restricted quantum computations $\mathbf{DQC_k}$, discussed in [39], known as the *one clean qubit model*, where the inputs to each circuit are restricted to be one pure qubit with as many maximally mixed qubits as desired. At the end of the computation, k qubits are measured in the computational basis. Clearly, $\mathbf{DQC_k} \subseteq \mathbf{BQP}$, but again, $\mathbf{DQC_k}$ is not believed to be universal for quantum computation.[6] It was shown in [39] that $\mathbf{PostDQC_k} = \mathbf{PP} = \mathbf{PostBQP}$ for $k \geq 3$. So, while $\mathbf{PostBQP} \subseteq \mathbf{PostDQC_k}$, under reasonable assumptions [40] it is not the case that $\mathbf{BQP} \subseteq \mathbf{DQC_k}$. Other examples demonstrating the same results are Boson Sampling [63], and any efficient computation generated by a two-qubit commuting Hamiltonian which generates entanglement [48].

2.4 Oracles

In classical computation, an *oracle* is a total function $O : \mathbb{N} \to \{0, 1\}$. A number x is said to be in an oracle O if $O(x) = 1$, hence oracles can decide membership in a language. Let $\mathbf{C}$ and $\mathbf{B}$ be complexity classes, then $\mathbf{C^B}$ denotes the class $\mathbf{C}$ with an oracle for $\mathbf{B}$. Informally, we can think of $\mathbf{C^B}$ as the class of languages decided by a computation which is subject to the restrictions and acceptance criteria of $\mathbf{C}$, but allowing an extra kind of computational step: an oracle for any desired language $\mathcal{L} \in \mathbf{B}$ that may be queried at any stage in the course of the computation, with each such query counting as a single computational step. That is, bit strings may be generated at any stage of the computation and presented to the oracle, which in a single step, returns the information of whether the bit string is in $\mathcal{L}$ or not. Given two complexity classes, $\mathbf{C}_1$ and $\mathbf{C}_2$, we say the relation[7] $\mathbf{C}_1 = \mathbf{C}_2$ holds relative to the oracle $\mathbf{B}$, if $\mathbf{C}_1^{\mathbf{B}} = \mathbf{C}_2^{\mathbf{B}}$. Such a result is referred to as a *relativised separation* result.

[6] In fact, under reasonable assumptions, $\mathbf{DQC_k}$ is provably not universal for quantum computation [40].

[7] The $=$ can be replaced with $\neq$, $\subseteq$ or $\supseteq$ equally well.

Oracles play a special role in quantum computation, forming the basis of most known computational speed ups over classical computation [1]. In quantum computation, oracle queries are represented by a family $\{R_n\}$ of quantum gates, one for each query length. Each R_n is a unitary transformation acting on $n+1$ qubits, whose effect on the computational basis is given by

$$R_n|x, a\rangle = |x, a \oplus A(x)\rangle$$

for all $x \in \{0, 1\}^m$ and $a \in \{0, 1\}$, where A is some Boolean function that represents the specific oracle under consideration. One could also consider more general oracles that, when queried, apply some general unitary transformation to the query state, but here, we only consider oracles that compute Boolean functions. In the state vector formalism of quantum theory, the action of a unitary oracle is defined on a maximal set of pure and perfectly distinguishable states, namely the computational basis. Linearly extending this to all states in the Hilbert space uniquely defines the action of the oracle on any state.

As pointed out to us by Howard Barnum [42], the situation for generalised probabilistic theories is more subtle. Consider, for example, the density matrix formulation of quantum theory, and suppose that oracle queries correspond to a family of trace-preserving completely-positive maps $\{\mathcal{E}_n\}$. Analogously to the state vector formalism, define the action of the oracle on a maximal set of pure and perfectly distinguishable states $\{\rho_i\}_{i=1}^N$, where each ρ_i is a density matrix, by

$$\mathcal{E}_n\big(\rho_x \otimes \rho_a\big) = \rho_x \otimes \rho_{a \oplus A(x)}, \tag{2.1}$$

where $\rho_x = \rho_{x_1} \otimes \cdots \otimes \rho_{x_n}$ and A is the function computed by the oracle. Note that

$$\rho_x \otimes \rho_a \to \rho_x \otimes \rho_{a \oplus A(x)} \quad \Longleftrightarrow \quad |x, a\rangle \to e^{i\phi(x,a)}|x, a \oplus A(x)\rangle,$$

where $a = 1, \ldots, N$ and $e^{i\phi(x,a)}$ is some phase factor that depends on the query state. Now, in addition to being able to compute the function A, a quantum computer with access to the oracle may also acquire information about the function ϕ, which may be hard to compute [43]. The usual definition of a quantum oracle therefore prevents 'sneaking in information' through phase factors.

In generalised probabilistic theories (with sufficient distinguishable states), it is easy to produce a definition of an oracle similar to that of equation 2.1. But for a system type A, a maximal set of pure and perfectly distinguishable states does not in general span the vector space $\mathbf{V_A}$. Hence the action of an oracle on such a set of states will not, in general, uniquely define its action on an arbitrary state in the state space. It is then not clear what extra condition must be placed on the oracle, first to define its action on arbitrary input states, and second to prevent non-trivial information being obtained through its action on non-basis input states (perhaps via a generalised notion of phase [130]).

Rather than attempt to solve this problem here (we show in chapter 6 that theories satisfying a certain number of physical principles do admit well-defined computational oracles, which reduce to the standard notion in the case of quantum theory), we will instead consider a notion of "classical oracle" that can be defined in any generalised probabilistic theory that satisfies the causality assumption of Section 1.3.2. The causality assumption allows the construction of adaptive circuits without paradox (see [15] for a more thorough discussion of the causality assumption, adaptive circuits, and conditioned transformations). In an adaptive circuit, the choice of which test to perform can depend on the outcomes $r_1, \ldots, r_k$ of previous tests in the circuit. An oracle $A : \mathbb{N} \to \{0, 1\}$ defines an extra gate that can be used in a computation in addition to those of the finite gate set, but with input and output that are classical wires, rather than being typed as with the gates intrinsic to the theory. The input to the oracle is a function $f(r_1, \ldots, r_k)$ of the outcomes of tests that appear in the circuit prior to the use of the oracle. The design of that portion of the circuit that is subsequent to the oracle can depend on the output $A(f)$ of the oracle. An oracle can be used in this way an unlimited number of times in a circuit, with each use counting as one gate. The uniformity condition must be extended, so that for each use of the oracle in a circuit, the input $f(r_1, \ldots, r_k)$, and the design of the circuit subsequent to the oracle, are computable in poly-time by a Turing machine with access to an oracle for $\mathbf{A}$. The acceptance criterion can also be extended so that for a circuit outcome z, the function $a(z)$ is computable in poly-time by a Turing machine with access to an oracle for $\mathbf{A}$.

Definition 2.9. *For each causal generalised probabilistic theory* $\mathbf{G}$*, a language* $\mathcal{L}$ *is in the class* $\mathbf{BGP}_{cl}^{\mathbf{A}}$ *if there exists a poly-size uniform family of circuits with access to the classical oracle* $\mathbf{A}$*, and an efficient acceptance condition, such that*

- $x \in \mathcal{L}$ *is accepted with probability at least* $\frac{2}{3}$.
- $x \notin \mathcal{L}$ *is accepted with probability at most* $\frac{1}{3}$

We can use the notion of classical oracle to obtain the following relativised separation result.

Theorem 2.10. *There exists a classical oracle* $\mathbf{A}$ *such that for any causal generalised probabilistic theory* $\mathbf{G}$, $\mathbf{NP^A} \not\subseteq \mathbf{BGP}_{cl}^{\mathbf{A}}$.

The proof is in section 2.5.4. This generalises the results of [45] from quantum theory to causal generalised probabilistic theories that satisfy tomographic locality. The result proved in section 2.5.4 is actually stronger: there exists a classical oracle $\mathbf{A}$ such that for any causal generalised probabilistic theory $\mathbf{G}$ that satisfies tomographic locality, the polynomial time hierarchy is infinite and $\mathbf{BGP}_{cl}^{\mathbf{A}} \subseteq \mathbf{P^A}$. The oracle in question is the same oracle that was used by Fortnow and Rogers in [45].

2.5 Proofs of the theorems

The proofs of each theorem presented in this chapter are provided in this section. Before we can give the proofs however, we must first discuss approximate circuit families.

2.5.1 Approximate circuit families

Consider a poly-size uniform circuit family $\{C_x\}$, defined over a finite gate set $\mathcal{G}$. Each gate in $\mathcal{G}$ corresponds to some finite set of transformations, one for each classical outcome of the gate. From the uniformity condition, the entries of the matrices representing these transformations can be calculated to accuracy ϵ in time $\mathrm{poly}(\log(1/\epsilon))$. With $\epsilon(|x|)$ a function of the input size, consider a family $\{\widetilde{C_x}\}$ of approximations to the original circuits, where matrix elements are replaced by rational numbers within $\epsilon(|x|)$ of the original matrix elements. Call $\{\widetilde{C_x}\}$ an $\epsilon(|x|)$-*approximation* to $\{C_x\}$. The following result shows that $\{\widetilde{C_x}\}$ can simulate $\{C_x\}$, to an accuracy dependent on $\epsilon(|x|)$.

Proposition 2.11. *Let* $\{C_x\}$ *be a uniform circuit family, with the number of gates in* C_x *bounded by a polynomial* $q(|x|)$. *Let* $\{\widetilde{C_x}\}$ *be an* $\epsilon(|x|)$*-approximation to* $\{C_x\}$, *with* $\epsilon(|x|) \leq 1$. *If the circuit* $C_T \in \{C_x\}$ *gives an outcome sequence* z *with probability* $P(z)$, *then the circuit* $\widetilde{C_T} \in \{\widetilde{C_x}\}$ *gives outcome sequence* z *with amplitude* $\widetilde{P}(z)$ *such that*

$$|P(z) - \widetilde{P}(z)| \leq D^{q(|T|)-1} q(|T|)\epsilon(|T|)N,$$

where N *and* D *are constants depending on the gate set* $\mathcal{G}$.

The word *amplitude* here should not be confused with the complex amplitudes of quantum theory. It is used for the real-valued quantity which approximates an outcome probability for the original circuit family, and is used rather than the term *probability*, because this quantity can be (slightly) less than 0 or (slightly) greater than 1. (The approximating circuit family is a mathematical construction that need not correspond precisely to a valid circuit family in the theory.) This proposition will be useful in the main proofs, since if $\{C_x\}$ is a circuit family that decides some language $\mathcal{L}$ in **BGP**, it follows that a $\frac{1}{12q(|x|)D^{q(|x|)-1}N}$-approximation to $\{C_x\}$ will accept a string $x \in \mathcal{L}$ with amplitude at least $7/12$, and will accept a string $x \notin \mathcal{L}$ with amplitude at most $5/12$, hence the success amplitude is still bounded away from $1/2$. The uniformity condition ensures that such an $\epsilon(|x|)$-approximation can be constructed in time polynomial in $|x|$.

In order to prove the proposition, two lemmas will be helpful.

Lemma 2.12. *Let M be a real $n \times m$ matrix such that for each entry, m_{ij}, we have that $|m_{ij}| \leq \epsilon$, for $\epsilon > 0$. Then*

$$\|M\|_{op} \leq nm\epsilon,$$

where $\|.\|_{op}$ is the operator norm.

Proof. Let M_i be the i^{th} row of M. Then

$$|M_i|_E = \sqrt{\sum_{j=1}^{m} m_{ij}^2} \leq \sum_{j=1}^{m} |m_{ij}| \leq \epsilon m,$$

where $|.|_E$ is the Euclidean norm, hence

$$|Mv|_E \leq \sum_{i=1}^{n} |M_i v| \leq \sum_{i=1}^{n} \epsilon m = nm\epsilon,$$

for $|v| = 1$, where the second inequality follows from the Cauchy-Schwarz inequality. Thus $\|M\|_{op} \leq nm\epsilon$. □

Lemma 2.13. *Let $\{M_i\}_{i=1}^{T}$ and $\{\widetilde{M_i}\}_{i=1}^{T}$ be two sets of matrices. Then the T-fold product of these matrices satisfies*

$$\|M_T \dots M_1 - \widetilde{M_T} \dots \widetilde{M_1}\|_{op} \leq D^{T-1} \sum_{i=1}^{T} \|M_i - \widetilde{M_i}\|_{op},$$

where $D = \max\{\|M_1\|_{op}, \dots, \|M_T\|_{op}, \|\widetilde{M_1}\|_{op}, \dots, \|\widetilde{M_T}\|_{op}\}$.

Proof. Consider the case of $T = 2$. With $|v| = 1$,

$$\begin{aligned}
&|\big(M_2M_1 - \widetilde{M_2}\widetilde{M_1}\big)v|_E \\
&= |\big(M_2M_1 - \widetilde{M_2}M_1\big)v + \big(\widetilde{M_2}M_1 - \widetilde{M_2}\widetilde{M_1}\big)v|_E \\
&\le |\big(M_2 - \widetilde{M_2}\big)M_1v|_E + |\widetilde{M_2}\big(M_1 - \widetilde{M_1}\big)v|_E \\
&\le \|M_2 - \widetilde{M_2}\|_{op}\|M_1\|_{op} + \|\widetilde{M_2}\|_{op}\|M_1 - \widetilde{M_1}\|_{op}.
\end{aligned}$$

Thus

$$\|M_2M_1 - \widetilde{M_2}\widetilde{M_1}\|_{op} \le D\|M_1 - \widetilde{M_1}\|_{op} + D\|M_2 - \widetilde{M_2}\|_{op}$$

The result follows from induction on T. □

We can now prove Proposition 2.11.

Proof. A particular outcome sequence of the circuit $C_T \in \{C_x\}$ corresponds to a sequence of matrices $\mathcal{G}^{r_1,1}, \dots, \mathcal{G}^{r_q,q}$, where $\mathcal{G}^{r_i,i}$ represents the r_ith outcome of the ith gate in C_T. Note that states and effects are included in this sequence. Tensoring these gates with identity transformations on systems on which they do not act and padding the corresponding matrices with rows and columns of zeros results in a sequence of square matrices $M^{r_q,q}, \dots, M^{r_1,1}$ such that

$$P(z) = P(r_1, \dots, r_q) = b^T.M^{r_q,q} \dots M^{r_1,1}.b,$$

where b is the vector $(1, 0, \dots, 0)$ and b^T is its transpose. Similarly for $\widetilde{\mathcal{G}}^{r_1,1}, \dots, \widetilde{\mathcal{G}}^{r_q,q}$, so that

$$\widetilde{P}(z) = \widetilde{P}(r_1, \dots, r_q) = b^T.\widetilde{M}^{r_q,q} \dots \widetilde{M}^{r_1,1}.b.$$

Note that $\|M^{r_i,i}\|_{op} \le \|\mathcal{G}^{r_i,i}\|_{op}$ and $\|\widetilde{M}^{r_i,i}\|_{op} \le \|\widetilde{\mathcal{G}}^{r_i,i}\|_{op}$, for all i. Therefore,

$$\begin{aligned}
|P(z) - \widetilde{P}(z)| &= |b^T\big(M^{r_q,q} \dots M^{r_1,1} - \widetilde{M}^{r_q,q} \dots \widetilde{M}^{r_1,1}\big)b| \\
&\le |b^T|_E|\big(M^{r_q,q} \dots M^{r_1,1} - \widetilde{M}^{r_q,q} \dots \widetilde{M}^{r_1,1}\big)b|_E \\
&\le D'^{q(|T|)-1}\sum_{n=1}^{q}\|M^{r_n,n} - \widetilde{M}^{r_n,n}\|_{op} \le D'^{q(|T|)-1}q(|T|)N\epsilon(|T|),
\end{aligned}$$

where if n_im_i is the size of the matrix $\mathcal{G}^{r_i,i}$, then

$$N = \max\{n_qm_q, \dots, n_1m_1\},$$

and

$$D' = \max\{\|\mathcal{G}^{r_1,1}\|_{op}, \dots, \|\mathcal{G}^{r_q,q}\|_{op}, \|\widetilde{\mathcal{G}}^{r_1,1}\|_{op}, \dots, \|\widetilde{\mathcal{G}}^{r_q,q}\|_{op}\}.$$

Note that, as circuits are built from finite gate sets, N is a constant. The first inequality follows from the Cauchy-Schwarz inequality, the second from that fact that $|b^T| = 1$ and lemma 2.13, the third from lemma 2.12, the fact that the sum has $q(|T|)$ entries and the fact that, as $\widetilde{C}_T$ is an ϵ-approximation of C_T, the matrix $M^{r_i,i} - \widetilde{M}^{r_i,i}$ has entries satisfying $|m_{ij} - \widetilde{m}_{ij}| \leq \epsilon$.

The reverse triangle inequality gives

$$\|\widetilde{\mathcal{G}}^{r_i,i}\|_{op} - \|\mathcal{G}^{r_i,i}\|_{op} \leq \|\widetilde{\mathcal{G}}^{r_i,i} - \mathcal{G}^{r_i,i}\|_{op} \leq N\epsilon(|T|).$$

With $\epsilon(|T|) \leq 1$, and

$$D'' = \max\{\|\mathcal{G}^{r_1,1}\|_{op}, \dots, \|\mathcal{G}^{r_q,q}\|_{op}\},$$

we have $D' \leq D \equiv D'' + N$, which completes the proof.

□

2.5.2 Proof of Theorem 2.2

One method of proving Theorem 2.2 is to use **GapP** functions. **GapP** functions were first studied in the context of quantum computation by Fortnow and Rogers in [45], where, among other things, they showed that **BQP** $\subseteq$ **AWPP**. A good discussion on **GapP** functions can be found in Watrous's survey of quantum complexity theory [52]. Proofs in this section are modifications and generalisations of proofs presented in [45, 52, 41].

Given a polynomial-time non-deterministic Turing machine M and input string x, denote by $M_{acc}(x)$ the number of accepting computation paths of M given input x, and by $M_{rej}(x)$ the number of rejecting computation paths of M given x. A function $f : \{0,1\}^* \to \mathbb{Z}$ is a **GapP** function if there exists a polynomial-time non-deterministic Turing machine M such that $f(x) = M_{acc}(x) - M_{rej}(x)$ for all input strings x.

Many complexity classes can be described in terms of **GapP** functions. For example the class **PP** can be defined as those languages $\mathcal{L}$ such that, for some **GapP** function f and any input string x, if $x \in \mathcal{L}$ then $f(x) > 0$ but if $x \notin \mathcal{L}$ then $f(x) \leq 0$. A useful class of **GapP** functions is provided by the following theorem.

Theorem 2.14. *Any function $f : \{0,1\}^* \to \mathbb{Z}$ that can be computed in poly-time by a Turing machine is a* **GapP** *function.*

For a proof, see [41, p.237].

The notation $\langle x, y\rangle$ denotes the pairing function [45]: that is, a poly-time computable function that maps the pair of strings x and y bijectively to the set of finite length strings $\{0,1\}^*$ such that, given $\langle x, y\rangle$, both x and y can be extracted in poly-time. The following proposition gives slight generalisations of standard closure properties of **GapP** functions.

Proposition 2.15. *For a polynomial q and* **GapP** *function f, let $h : \{0,1\}^* \to \mathbb{Z}$ be defined for all $x \in \{0,1\}^*$ by*

$$h(x) = \sum_{\substack{|y| \leq q(|x|) \\ y \in L_x}} f(\langle x, y\rangle),$$

where L_x is some set (that may depend on x) with the property that membership of y in L_x can be determined in time polynomial in $|x|$. Then h is a **GapP** *function.*

Now let $g : \{0,1\}^* \to \mathbb{Z}$ be defined for all $x \in \{0,1\}^*$ by

$$g(x) = \prod_{\substack{1 \leq i \leq q(|x|) \\ i \in L_x}} f(\langle x, i\rangle),$$

where the symbol i appearing as the second argument on the pairing is a binary encoding of i and L_x is some set with the property that membership of i in L_x can be determined in time polynomial in $|x|$. Then g is also a **GapP** *function.*

Proof. We will prove the first statement only as the second statement follows from a similar generalisation of a standard argument. Let $f(x) = M_{acc}(x) - M_{rej}(x)$ for some non-deterministic poly-time Turing machine, M. Let N be a non-deterministic poly-time Turing machine that, on input $x \in \{0,1\}^*$, guesses a string y of length $\leq q(|x|)$, decides whether y is in L_x, and

- if $y \in L_x$, simulates M on input $\langle x, y\rangle$.
- if $y \notin L_x$, guesses a bit b and accepts if and only if $b = 0$.

N runs in poly-time, and for every $x \in \{0,1\}^*$, $N_{acc}(x) - N_{rej}(x) = h(x)$, hence h is a **GapP** function. □

For the rest of this section, assume that the pairing function is used whenever a function has two or more arguments. **GapP** functions are intimately related to computation in generalised probabilistic theories, as the following result shows.

Theorem 2.16. *Let $\{C_x\}$ be a poly-size uniform family of circuits in a generalised probabilistic theory. Then for any polynomial w and constant D, there exists a function $\epsilon(|x|) \leq 1/D^{w(|x|)}$, and an $\epsilon(|x|)$-approximation $\{\widetilde{C_x}\}$ to $\{C_x\}$, such that the amplitude for acceptance*[8] *of a circuit $\widetilde{C_T} \in \{\widetilde{C_x}\}$ is given by*

$$\widetilde{P}_T(\text{accept}) = \frac{f(T)}{2^{p(|T|)}},$$

where f is a **GapP** *function and $p(|T|)$ is a polynomial in the size of the input string.*

Proof. It follows from the uniformity condition that for any polynomial w, there is an $\epsilon(|x|)$-approximation $\{\widetilde{C_x}\}$ to $\{C_x\}$, with $\epsilon(|x|) \leq 1/D^{w(|x|)}$, such that the entries in the matrices representing gates in the circuit $\widetilde{C_T} \in \{\widetilde{C_x}\}$ have rational entries, and can be computed in time polynomial in $|T|$. Furthermore, the rational entries can be taken to have the form $c/2^d$, with $c \in \mathbb{Z}$, $d \in \mathbb{N}$, and d a polynomial function of $|T|$. Padding circuits with identity gates if necessary, assume that the number of gates in the circuit $\widetilde{C_T}$ is given by a polynomial function $q(|T|)$. A particular outcome of the circuit corresponds to matrices $\widetilde{\mathcal{G}}^{r_1,1}, \dots, \widetilde{\mathcal{G}}^{r_q,q}$, where $\widetilde{\mathcal{G}}^{r_i,i}$ represents the r_ith outcome of the ith gate in $\widetilde{C_T}$. States and effects are included in this sequence.

By tensoring these gates with identity transformations on systems on which they do not act and padding the corresponding matrices with rows and columns of zeros, we can obtain a sequence of square matrices $\widetilde{\mathcal{M}}^{r_1,1}, \dots, \widetilde{\mathcal{M}}^{r_q,q}$, such that (i) rows and columns of these matrices are indexed by bit strings of length $y(|T|)$, with $y(|T|)$ a polynomial function, and (ii) the amplitude of outcome $z = r_1, \dots, r_q$ is given by

$$b^T . \widetilde{\mathcal{M}}^{r_q,q} \cdots \widetilde{\mathcal{M}}^{r_1,1} .b,$$

where b is the vector $(1, 0, \dots, 0)$ and b^T is its transpose. Note that for each $\widetilde{\mathcal{M}}^{r_i,i}$, the matrix $2^d \widetilde{\mathcal{M}}^{r_i,i}$ has integer entries.

Consider the function $h : \{0,1\}^* \to \mathbb{Z}$ given by

$$h(T, r_1, \dots, r_q, n, i_0, \dots, i_q) = M^{r_n,n}_{i_n i_{n-1}},$$

where $i_0, \dots, i_q$ are bit strings of length $y(|T|)$, and $M^{r_n,n}_{i_n i_{n-1}}$ is the $i_n i_{n-1}$ entry of the matrix $2^d \widetilde{\mathcal{M}}^{r_n,n}$. By the uniformity condition, these matrix entries can be calculated in polynomial time by a Turing machine, so by Theorem 2.14, h is a **GapP** function.

[8] Note that, as $\{\widetilde{C_x}\}$ is a mathematical construction, it need not correspond to a valid circuit family in the theory and so cannot be said to accept or reject an input string. However, for ease of notation, we will say an approximating circuit 'accepts' an input string if $a(z) = 0$ where z is the outcome sequence of that approximating circuit, and 'rejects' the input string otherwise.

The amplitude for outcome $z = r_1 \dots r_q$ is given by

$$\begin{aligned}
\widetilde{P}(z) &= \frac{1}{2^{dq}} \sum_{\{i_1,\dots,i_{q-1}\}} M^{r_q,q}_{1i_{q-1}} \dots M^{r_2,2}_{i_2 i_1} M^{r_1,1}_{i_1 1} \\
&= \frac{1}{2^{dq}} \sum_{\{i_1,\dots,i_{q-1}\}} \prod_{1\le n\le q} h(T, r_1, \dots, r_q, n, i_0 = 1, i_1, \dots, i_{q-1}, i_q = 1) \\
&= \frac{1}{2^{dq}} \sum_{\{i_1,\dots,i_{q-1}\}} g(T, r_1, \dots, r_q, i_1, \dots, i_{q-1}), \\
&= \frac{f'(T,z)}{2^{dq}},
\end{aligned}$$

where g is a **GapP** function by Proposition 2.15, hence f' is a **GapP** function by another application of Proposition 2.15.

The amplitude for the circuit $\widetilde{C_T}$ to accept is given by

$$\widetilde{P}_T(\text{accept}) = \sum_{a(z)=0} \widetilde{P}_T(z) = \sum_{a(z)=0} \frac{f'(T,z)}{2^{dq}},$$

where $a(z)$ is the function that determines if z is an accepting or rejecting outcome. By the uniformity condition, $a(z)$ can be calculated in polynomial time by a Turing machine, hence Proposition 2.15 gives

$$\widetilde{P}_T(\text{accept}) = \frac{f(T)}{2^{p(|T|)}},$$

where f is a **GapP** function and $d(|T|)q(|T|) = p(|T|)$ is a polynomial that takes values in $\mathbb{N}$. □

Recall that the class **AWPP** (Almost Wide Probabilistic Polynomial time) can be defined [53] as follows.

Definition 2.17. *The class* **AWPP** *consists of those languages $\mathcal{L}$ such that there exists a* **GapP** *function f, and a polynomial r such that*

- *If $x \in \mathcal{L}$ then $2/3 \le f(x)/2^{r(|x|)} \le 1$,*
- *if $x \notin \mathcal{L}$ then $0 \le f(x)/2^{r(|x|)} \le 1/3$.*

The $1/3 - 2/3$ separation can be replaced by any constant, positive, separation [53].

Theorem 2.18. *For any generalised probabilistic theory* **G***,* **BGP** $\subseteq$ **AWPP***.*

Proof. If a language $\mathcal{L} \in \mathbf{BGP}$, then there is a poly-size uniform circuit family $\{C_x\}$ such that $P_x(\text{accept}) \geq 2/3$ if $x \in \mathcal{L}$, and $P_x(\text{accept}) \leq 1/3$ if $x \notin \mathcal{L}$. Assume that for all x, $1/10 \leq P_x(\text{accept}) \leq 9/10$.[9] By Theorem 2.16, there is an $\epsilon(|x|)$-approximation to $\{C_x\}$ such that the amplitudes determined by the approximating family satisfy

$$\widetilde{P}_x(\text{accept}) = \frac{f(x)}{2^{p(|x|)}},$$

with f a **GapP** function. Furthermore, for any polynomial w, $\epsilon(|x|)$ can be chosen so that $\epsilon(|x|) \leq 1/D^{w(|x|)}$. Hence by Proposition 2.11, $\epsilon(|x|)$ can be chosen small enough that $\widetilde{P}_x(\text{accept}) \geq 7/12$ if $x \in \mathcal{L}$ and $\widetilde{P}_x(\text{accept}) \leq 5/12$ if $x \notin \mathcal{L}$, and for all x, $0 \leq \widetilde{P}_x(\text{accept}) \leq 1$. Taking $p(|x|)$ to be the function $r(|x|)$ in definition 2.3 and noting that $5/12 - 7/12$ is a constant, positive, separation, gives the result. □

It is well known that $\mathbf{AWPP} \subseteq \mathbf{PP} \subseteq \mathbf{PSPACE}$ (see, for example, [55] and references therein).

2.5.3 Proof of Theorem 2.7

An alternate definition of the class **PP** can be stated [56, 53] as follows.

Definition 2.19. *The class* **PP** *consists of those languages* $\mathcal{L}$ *such that there exist* **GapP** *functions* f *and* h *so that for all* x

- *If* $x \in \mathcal{L}$ *then* $2/3 \leq f(x)/h(x) \leq 1$,
- *if* $x \notin \mathcal{L}$ *then* $0 \leq f(x)/h(x) \leq 1/3$.

The $1/3 - 2/3$ separation can be replaced by any constant, positive, separation [53].

In order to prove Theorem 2.7, consider a uniform family of circuits $\{C_x\}$ in the generalised probabilistic theory $\mathbf{G}$. Let S_T be a subset of the possible outcomes of the circuit C_T, with respect to which post-selection is defined, so that $P_T(\text{accept}|S_T) \geq 2/3$ for $T \in \mathcal{L}$ and $\leq 1/3$ for $T \notin \mathcal{L}$. As in the proof of Theorem 2.2, assume that

[9] This can be ensured, if necessary, by considering the circuit C_T to be carried out in parallel with a biased coin toss. With probability 1/5, the coin is tails, in which case the output of the circuit is ignored, and acceptance/rejection are returned with probability 1/2 each. Taken together, these circuits and coin tosses define a modified circuit family $\{C'_x\}$, and in the following, approximating circuit families can be assumed to be defined relative to $\{C'_x\}$.

these probabilities are also bounded away from 0 and 1 so that for all T, $1/10 \leq P_T(\text{accept}|S_T) \leq 9/10$.[10]

By Theorem 2.16, there is an $\epsilon(|x|)$-approximation to $\{C_x\}$ such that, in the approximating family, the joint amplitude to accept the computation and have an outcome from the set S_T is

$$\widetilde{P}_T(\text{accept}, S_T) = \frac{f(T)}{2^{p(|T|)}},$$

with f a **GapP** function. Similarly,

$$\widetilde{P}_T(S_T) = \frac{g(T)}{2^{q(|T|)}},$$

with g a **GapP** function and q a polynomial. Furthermore, for any polynomial w and constant D, $\epsilon(|x|)$ can be chosen so that $\epsilon(|x|) \leq 1/D^{w(|x|)}$. Hence by Proposition 2.11 and the fact that we are post-selecting on at most exponentially-unlikely outcomes, $\epsilon(|x|)$ can be chosen small enough that for the approximating circuit family, $\widetilde{P}_T(S_T) > 0$. This means that for the approximating circuit family, the conditional

$$\widetilde{P}_T(\text{accept}|S_T) = \frac{\widetilde{P}_T(\text{accept}, S_T)}{\widetilde{P}_T(S_T)},$$

is well defined. Furthermore, $\epsilon(|x|)$ can be chosen small enough that $\widetilde{P}_T(\text{accept}|S_T) \geq 7/12$ if $x \in \mathcal{L}$, $\widetilde{P}_T(\text{accept}|S_T) \leq 5/12$ if $x \notin \mathcal{L}$, and using the assumption that the original circuit family probabilities are bounded away from 0 and 1, the approximating amplitudes satisfy $0 \leq \widetilde{P}_T(\text{accept}|S_T) \leq 1$.

Now,

$$\widetilde{P}_T(\text{accept}|S_T) = \frac{2^{q(|T|)}f(T)}{2^{p(|T|)}g(T)} = \frac{l(T)}{h(T)},$$

where $h(T) = 2^{p(|T|)}g(T)$ and $l(x) = 2^{q(|T|)}f(T)$ are **GapP** functions. This follows from Theorem 2.14, Proposition 2.15, and the fact that both p and q are polynomials taking values in $\mathbb{N}$. The result follows.

2.5.4 Proof of Theorem 2.10

Denote by **PH** the polynomial time hierarchy: the union of an infinite hierarchy of classes $\mathbf{\Sigma}_k$, $\mathbf{\Delta}_k$ and $\mathbf{\Pi}_k$ for $k \in \mathbb{N}$, where $\mathbf{\Sigma}_0 = \mathbf{\Delta}_0 = \mathbf{\Pi}_0 = \mathbf{P}$ and $\mathbf{\Sigma}_{k+1} = \mathbf{NP}^{\mathbf{\Sigma}_k}$,

[10] This can be done, as before, by the introduction of a biased coin parallel to the circuit. If the circuit outcome is in S_T and the coin is heads, then accept or reject, depending on the circuit outcome. If the outcome is in S_T and the coin is tails then accept or reject with probability 1/2 each.

$\mathbf{\Delta}_{k+1} = \mathbf{P}^{\mathbf{\Sigma}_k}$ and $\mathbf{\Pi}_{k+1} = \mathbf{coNP}^{\mathbf{\Sigma}_k}$. The polynomial time hierarchy is a natural way of classifying the complexity of problems beyond the class **NP**. It is a strongly held belief in computer science that **NP** includes non-polynomial-time problems.

Theorem 2.10 is a corollary of two results, the first of which is due to [55] and [57]:

Theorem 2.20. *There exists an oracle* $\mathbf{A}$ *such that* $\mathbf{P}^{\mathbf{A}} = \mathbf{AWPP}^{\mathbf{A}}$ *and the polynomial time hierarchy is infinite.*

The second is that Theorem 2.18 *relativizes.*

Theorem 2.21. *For any classical oracle* $\mathbf{A}$ *we have that* $\mathbf{BGP}_{cl}^{\mathbf{A}} \subseteq \mathbf{AWPP}^{\mathbf{A}}$ *for any causal* $\mathbf{G}$.

Proof. Given the uniformity condition for circuit families with an oracle, entries in the matrices representing gates in a circuit are all computable in polynomial time by a Turing machine with access to the oracle $\mathbf{A}$. Thus the proof of Theorem 2.16 goes through essentially unchanged, except that in this case the conclusion is that the acceptance amplitude is

$$\widetilde{P}_x(\text{accept}) = \frac{f(x)}{2^{p(|x|)}},$$

where $p(|x|)$ is a polynomial function of the size of the input and f is a $\mathbf{GapP}^{\mathbf{A}}$ function. A $\mathbf{GapP}^{\mathbf{A}}$ function is defined in a similar fashion to a **GapP** function, except instead of counting the difference between the number of accepting and rejecting paths for any input into a non-deterministic Turing machine, $\mathbf{GapP}^{\mathbf{A}}$ functions count the difference between the number of accepting and rejecting paths for any input into a non-deterministic Turing machine with access to the oracle $\mathbf{A}$.

$\mathbf{AWPP}^{\mathbf{A}}$ can be defined with respect to $\mathbf{GapP}^{\mathbf{A}}$ functions by just replacing every mention of **GapP** functions with $\mathbf{GapP}^{\mathbf{A}}$ functions in Definition 2.3. Thus the proof that $\mathbf{BGP}_{cl}^{\mathbf{A}} \subseteq \mathbf{AWPP}^{\mathbf{A}}$, for any causal GPT and oracle $\mathbf{A}$, goes through exactly the same as the proof of Theorem 2.18. □

Hence we obtain

Theorem 2.22. *There exists a classical oracle* $\mathbf{A}$ *relative to which* $\mathbf{BGP}_{cl}^{\mathbf{A}} \subseteq \mathbf{P}^{\mathbf{A}}$, *for all causal* $\mathbf{G}$, *and the polynomial time hierarchy is infinite.*

This implies that there exists a classical oracle relative to which **NP** is not contained in **BGP**, for any causal theory **G** satisfying tomographic locality. This generalises the results of [45] from quantum theory to general theories.

2.6 Discussion and conclusion

In this chapter we have begun an investigation into the relationship between computation and physical principles. Using the circuit framework approach to generalised probabilistic theories, we investigated the computational power of theories formulated in operational terms, along with the role played by simple physical principles. We started by defining a rigorous model of computation which allowed for the definition of the complexity class of problems efficiently solvable by a specific theory. The strongest known inclusion for the quantum case, **BQP** $\subseteq$ **AWPP**—which implies **BQP** $\subseteq$ **PP** $\subseteq$ **PSPACE**—was shown to still hold in any theory satisfying tomographic locality. It is notable that this includes even those theories that violate the causality principle. One possible interpretation of this is that we should *in principle* be able to derive stronger upper bounds for the quantum class **BQP** as this result illuminates the fact that the "quantum" proof of this upper bound does not exploit any of the structure unique to quantum theory. Combining these results with a result of Aaronson's, it follows that any problem efficiently solvable in a theory satisfying tomographic locality can also be solved efficiently by a post-selecting quantum computer. In fact, one can say something stronger: any problem efficiently solvable with post-selection in a theory satisfying tomographic locality can also be solved efficiently by a post-selecting quantum computer. Roughly speaking, then, in a world with post-selection, quantum theory is optimal for computation in the space of all tomographically local theories.

We discussed the problem of defining a computational oracle for an arbitrary theory. In general, this problem may have no general solution if it is required that the definition of an oracle reduce to the standard definition in the quantum case. Nonetheless, a notion of "classical oracle" was defined in any theory that satisfies the causality principle, and for such theories there exists a classical oracle relative to which **NP** is not contained in **BGP**. However, we show in chapter 6 that there is an interesting subclass of theories—satisfying causality, purification, purity preservation, and strong symmetry—for which a notion of oracle can be defined that admits 'superposition' of inputs, and reduces to the standard definition in the quantum case. In such theories, the solution of the 'subroutine problem' of [44] might serve as an interesting computational principle that could rule out certain theories, potentially providing a new principle from which quantum theory can be derived. We discuss this point further in chapter 6.

It would be interesting to determine whether violation of the causality principle can confer extra computational power. An initial thought is that there could exist a non-causal theory that can efficiently solve **NP**-complete problems. Given that the inclusion **BGP** $\subseteq$ **AWPP** holds even for non-causal—but tomographically local—theories, this can only be the case if **NP** is contained in **AWPP**. At present, this is unknown, and establishing the question either way would constitute a major advance in complexity theory [46, 53]. This is due to the fact that, intuitively, the inclusion **NP** $\subseteq$ **AWPP** implies one has very strong control over the possible number of accepting computations of **NP** machines [46, 53]. Nevertheless, it would be interesting if a violation of causality enabled the efficient solution of other problems, thought to be hard for quantum computers, but known to be in **AWPP**.

Moreoever, although the main results of this chapter do not require the causality principle, we have nonetheless been considering circuits in which gates appear in a fixed structure. It would be interesting to investigate the computational power of theories in which there is no such definite structure. Frameworks for describing situations with indefinite causal structure have been defined with the aim of discussing aspects of quantum gravity [50, 51]. Some preliminary remarks on the computational power of such theories were given in [61]. A specific query complexity problem that can be solved with fewer queries on a quantum computer in which the gates do not appear in a fixed order than on a standard quantum computer was presented in [62].

An open question is to establish tighter bounds on the power of general theories. Even with tomographic locality assumed, there is a lot of freedom in the construction of a generalised theory. Is there an explicit construction that solves a hard problem, that is, a problem at least thought to be hard for quantum computers? Even better, can we describe a complexity class, potentially larger than **BQP**, and an explicit construction of a general theory **G**, such that this class is contained in **BGP**? This question is the main focus of the next chapter.

Chapter 3

The computational landscape of generalised theories

In chapter 2, we defined and studied a circuit based model of computation in the operationally-defined framework of generalised probabilistic theories. One of the main results of that chapter was to show that, for any theory satisfying tomographic locality—whether or not causality is satisfied—computational problems that can be solved efficiently are contained in the classical complexity class **AWPP**, a fact first proved in the quantum case by Fortnow and Rogers [45]. A natural question then arises: does there exist a tomographically local theory **G** such that **BGP** = **AWPP**? Such a theory could be used as a "foil" to deepen our understanding of the limitations of quantum computers, in much the same way as PR boxes deepened our understanding of quantum non-locality.

In section 3.2 of this chapter, we present a complexity-theoretic argument which suggests that such a possibility may be unlikely. We show that if

$$\mathbf{PromiseBGP} = \mathbf{PromiseAWPP}$$

for a tomographically local theory **G**, then one necessarily has $\mathbf{NP} \subseteq \mathbf{AWPP}$. Here, **PromiseBGP** and **PromiseAWPP** are *promise* versions of the classes **BGP** and **AWPP**, meaning that they contain *promise* rather than decision[1] problems. Promise problems only demand a solution in the situation where the problem input satisfies a certain condition, known as the promise. While **PromiseBGP** = **PromiseAWPP** may hold independently of whether **BGP** = **AWPP** holds, these statements are at least conceptually related. Intuitively, if one of these statements appears unlikely, the other one should also be considered unlikely—although to a lesser degree. As discussed at the end of chapter 2, it is believed unlikely [53, 46] that **NP** is contained

[1] An example of such a problem is deciding whether some bit string is in a particular language.

in **AWPP**. This can, in some sense, be taken as evidence against the existence of a theory whose computational power (in the promise problem setting) exactly equals **PromiseAWPP**. Which, as briefly argued above, may constitute evidence (albeit of a much weaker variety) against the existence of a theory whose computational power exactly equals **AWPP**.

In this chapter we present a slight relaxation of the standard definition of generalised probabilistic theories, which was originally introduced in chapter 1, and show that—despite all of the above—one can construct a theory within this altered framework, satisfying both tomographic locality and causality, in which the class of efficiently solvable problems exactly equals **AWPP**, but such that this result does not immediately imply **NP** $\subseteq$ **AWPP**. Hence **AWPP**, despite having a slightly involved definition in terms of gap functions for non-deterministic Turing machines, can be thought of much more intuitively as the class of problems efficiently solvable by tomographically local physical theories.

In the standard definition of a generalised theory—presented in chapter 1—a theory specifies a set of laboratory devices from which one can construct closed circuits by composing devices in sequence and parallel, and assigns a probability distribution over the outcomes of each closed circuit. Moreover, the set of devices—and device outcomes—is closed under sequential and parallel composition. For the purposes of this chapter, we refer to such theories as "free" generalised probabilistic theories. In the modified definition of a generalised probabilistic theory, which shall be introduced in section 3.1, a theory specifies a set of devices, a set of *allowed* closed circuits which can be built from those devices, and assigns a probability distribution over the outcomes of allowed closed circuits. Note that this modified definition is slightly more general than the one introduced in chapter 1 as it only assigns a probability distribution to the set of *allowed* closed circuits which are specified by the theory. However, our definition is not unmotivated if one takes the viewpoint that a physical theory corresponds both to a consistent account of experimental data and to which experiments are implementable in principle.

Note that the modified definition of generalised probabilistic theories will only be important for this chapter. In all other chapters—unless explicitly stated otherwise—the standard framework introduced in chapter 1 is used.

3.1 The modified framework

In this section the weakened definition of generalised probabilistic theories will be presented. We start by recalling some of the key terminology of chapter 1 and then provide a rigorous definition of a theory within this new framework, all the while highlighting the differences (and similarities) to the standard definition presented in chapter 1. We show how to formulate finite tomography, local tomography, and causality in this new framework, and demonstrate that—despite the weakened constraints on what constitutes a theory—in all theories satisfying both finite and local tomography each set of transformations still gives rise to a finite dimensional vector space structure in which parallel composition corresponds to the standard vector space tensor product. Note again that the modified definition of generalised probabilistic theories will only be important for this chapter. In all other chapters—unless explicitly stated otherwise—the standard framework introduced in chapter 1 is used.

As discussed in chapter 1, a laboratory device comes equipped with input ports, output ports, and a classical pointer. When a device is used in an experiment, the pointer comes to rest in one of a number of positions, indicating that a particular outcome has occurred. Input and output ports are *typed*, with types given by labels $A, B, C, \ldots$. Closed circuits are built by composing[2] devices in sequence and in parallel.

More formally, a theory specifies:

1. A set of devices.

2. An allowed set of closed circuits built by composing those devices in sequence and parallel.

3. An assignment of a probability distribution over the outcomes of each allowed closed circuit, such that each closed circuit corresponds to a collection $\{p_i\}_{i \in Z}$ were Z is the outcome set of the circuit, $p_i \in [0,1]$ and $\sum_i p_i = 1$.

All of the above must satisfy the following two constraints: (i) the set of closed circuits is closed under parallel composition, and (ii) if a circuit consists of two, or more, separate closed circuits in parallel, then the overall probability for a particular fixed outcome on each closed circuit factors over the constituent closed circuits (as was the case in chapter 1). Note that in a specific theory, the set of devices may not

[2] The definition of sequential and parallel composition employed here is the same as the one presented in chapter 1.

be closed under sequential and parallel composition. Only those compositions which result in closed circuits allowed in the theory are possible.

To each device $\mathcal{E}$ there corresponds a set $C_{\mathcal{E}}$ of *complement circuits*, that is, appropriately typed circuit fragments into which one can insert $\mathcal{E}$ to get a correctly typed closed circuit, denoted $C(\mathcal{E})$ for each $C \in C_{\mathcal{E}}$. Note that $C(\mathcal{E})$ may not be an allowed closed circuit in the theory, hence $C_{\mathcal{E}}$ is the set of mathematically conceivable—correctly typed—closed circuits built with G. If $C(\mathcal{E})$ is an allowed closed circuit in the theory, denote by $C_m(\mathcal{E}_k)$ the probability of observing outcome m on the fragment C and outcome k on the device $\mathcal{E}$ when the closed circuit $C(\mathcal{E})$ is constructed. If $C(\mathcal{E})$ is not allowed by the theory, then all closed circuit outcomes $C_m(G_k)$ are assigned the value 'u' for "undefined". Two device outcomes are operationally equivalent from the point of view of a fixed complement circuit outcome if either both are defined and give the same probability, or both of the closed circuit outcomes are undefined.

Equivalence classes of preparation outcomes defined in this way will be called *states* and equivalence classes of measurement outcomes will be called *effects.* Equivalence classes of general device outcomes, where the devices have both a non-zero number of input ports and a non-zero number of output ports, are called *transformations.* As in chapter 1, the set of transformations with input port of type A and output port of type B is denoted $\mathbf{Transf}(\mathbf{A},\mathbf{B})$, the set of states with output port of type B is denoted $\mathbf{St}(\mathbf{B})$ and the set of effects of input port of type A is denoted $\mathbf{Eff}(\mathbf{A})$.

Finite tomography. *For each set* $\mathbf{Transf}(\mathbf{A},\mathbf{B})$*, there exist a finite and minimal set of appropriately typed circuit outcome fragments* $\{f_i\}_{i=1}^n$ *such that: (i) all closed circuits that can be built with them are allowed in the theory, and (ii) for any* $T \in \mathbf{Transf}(\mathbf{A},\mathbf{B})$ *and any real linear combination of transformations from* $\mathbf{Transf}(\mathbf{A},\mathbf{B})$*,* $\sum_k \gamma_k \mathcal{F}^k$*, if*

$$f_i(T) = f_i\left(\sum_k \gamma_k \mathcal{F}^k\right), \quad \forall i = 1, \ldots, n,$$

then we have $T = \sum_k \gamma_k \mathcal{F}^k$*.*

We now show that any theory which satisfies the principle of finite tomography admits a finite vector space for each type. Indeed, note that each transformation T defines a function $\widetilde{T}$ from the set of complement circuits C_T to $[0,1] \cup \{u\}$. It is assumed in the definition of finite tomography that any closed circuit built with the f_i's is allowed in the theory. Hence each transformation T of the correct type defines a function $\widehat{T}$ from the set $\{f_i\}_{i=1}^n$ to $[0,1] \subset \mathbb{R}$, satisfying $\widehat{T} = \widetilde{T}|_{\{f_i\}}$. We shall now use this fact to extend the function $\widetilde{T}$ to a fully defined function from C_T to the real numbers, i.e. a function which no longer takes the value 'u'. Choose a set of n transformations $\{T^i\}_i$ with the property that for any T^i, there does not exist any real coefficients λ_j such that $f_k(T^i) = f_k\left(\sum_{j\neq i} \lambda_j T^j\right) = \sum_{j\neq i} \lambda_j f_k(T^j)$ for all f_k. That is, no real linear combination of any $n-1$ equals the remaining one. Note that, as discussed in chapter 1, such a set must exist due to the minimality of the f_i's. Now, for any compliment circuit outcome C, if $C(T^i) = u$ for any i, then assign it an arbitrary (but fixed once the choice is made) real value. In this manner, we extend the partially defined real-valued function $\widehat{T^i}$ to a totally defined real-valued function $\widehat{T^i}'$ such that $C(\widehat{T^i}') = C(\widehat{T^i})$ whenever $C(\widehat{T^i}) \neq u$, and $C(\widehat{T^i}')$ takes the newly assigned real number whenever $C(T^i) = u$. One can take real linear combinations of such functions $\widehat{h} = \sum_i \lambda_i \widehat{T^i}'$ whose action on complement circuits is defined via linear extension $C(\widehat{h}) := \sum_i \lambda_i C(\hat{T^i})$. By taking the linear span of $\{T^i\}$, one generates a finite dimensional real vector space. Note that, by construction, the vectors $\{T^i\}$ are linearly independent in this vector space. Using the argument from section 1.2 of chapter 1, it follows that for each transformation T there exists a set of real numbers $\{\alpha_i\}$ such that $T = \sum_i \alpha_i T^i$. Hence, the set of transformations $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$ belongs to this vector space. Moreover, we have shown that each transformation T gives rise to a function from appropriately typed circuit fragments to the real numbers.

As depicted in diagram 1.4 from chapter 1, a circuit outcome fragment which corresponds to attaching a state to each input port and an effect to each output port of a transformation which, when inserted into the circuit outcome fragment forms a closed circuit, will be referred to as *local circuit fragments.*

Tomographic locality. *If, for any transformation $T \in$ $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$ and any real linear combination of transformations $\sum_k \gamma_k \mathcal{F}^k$,*

$$l(T) = l\left(\sum_k \gamma_k \mathcal{F}^k\right) = \sum_k \gamma_k l\left(\mathcal{F}^k\right)$$

for all local circuit outcome fragments l, then we have that $T = \sum_k \gamma_k \mathcal{F}^k$.

A consequence of tomographic locality is that for a transformation with input type AB and output type CD, the corresponding real vector space decomposes into a tensor product of vector spaces as in equation 1.5 from chapter 1. This follows from the fact that local circuit outcome fragments are separating (recall the difference between separating for the spanning set versus separating for vector space spanned by that set discussed in chapter 1) for the vector space generated by the set of transformations. Furthermore, a transformation $T \in \mathbf{Transf}(\mathbf{A}, \mathbf{B})$ is completely specified by its action on $\mathbf{St}(\mathbf{A})$, hence T can be identified with a unique linear map acting on the vector space generated by the set of states.

Causality. *For each type there exists a unique deterministic effect U, with the property that any closed circuit built with U is allowed in the theory, such that $\sum_r e_r = U$ for all appropriately typed measurements $\{e_r\}_r$.*

Mathematically, causality is equivalent to the statement: "Probabilities of present experiments are independent of future measurement choices".

The standard definition of a theory presented in chapter 1—which we refer to here as a *free theory*—is presented below.

Free theory. *A* free *theory is specified by a collection of devices which are closed under parallel and sequential composition, such that each closed circuit corresponds to a collection $\{p_i\}_{i \in Z}$ where Z is the outcome set of the circuit, $p_i \in [0, 1]$ and $\sum_i p_i = 1$. Moreover, probabilities for independent circuits factorise.*

It is clear that free theories are a special case of the more general theories considered in this chapter. As devices in free theories are closed under both parallel and sequential composition, any closed circuit built with these devices is allowed in the theory.

Note that theorem 2.2 from chapter 2, which proved **AWPP** upper bounded the class of problems efficiently solvable in any tomographically local (free) theory, can be straightforwardly shown to hold for the non-free theories satisfying tomographic locality introduced here.

3.2 Evidence against free theories achieving the upper bound

We now present a complexity-theoretic argument which can, in some sense, be considered weak evidence (recall the discussion at the start of this chapter) against the existence of a free theory whose computational power exactly equals **AWPP**.

Theorem 3.1. *If there exists a tomographically local free theory $\boldsymbol{G}$ satisfying*

$$\mathbf{PromiseBGP} = \mathbf{PromiseAWPP},$$

then one necessarily has

$$\mathbf{NP} \subseteq \mathbf{AWPP}.$$

As discussed at the start of this chapter, **PromiseBGP** and **PromiseAWPP** are *promise* versions of the classes **BGP** and **AWPP**, meaning that they contain *promise* rather than decision problems. An example of a decision problem is deciding whether a certain bit string is contained in a particular language. A promise problem is a generalisation of a decision problem where the input is *promised* to belong to a subset of all possible inputs. Unlike decision problems, the accepting instances (the inputs for which an algorithm must accept) and the rejecting instances do not exhaust the set of all inputs. There may be inputs which are neither accepting or rejecting. If such an input is given to an algorithm for a certain promise problem, no requirements are placed on the output, i.e. the algorithm is allowed to output anything.

Proof. Firstly, it is known that the Unique Satisfiability Problem (the problem of deciding whether a given Boolean formula has exactly one satisfying truth assignment, or no satisfying assignment at all, promised that one of these is the case) is contained in **PromiseAWPP** [54]. The Valiant-Vazirani theorem [47] says that if one has an efficient algorithm for solving the Unique Satisfaction Problem in conjunction with the ability to probabilistically amplify polynomially-unlikely events, then one can solve any problem in **NP**. More formally, the Valiant-Vazirani theorem says the standard Boolean Satisfiability Problem can be randomly reduced to the Unique Satisfiability Problem. Now, if **PromiseBGP** = **PromiseAWPP** then the Unique Satisfaction Problem is in **PromiseBGP**. That is, there exists an algorithm (i.e. a specific circuit family) in **G** which, when given inputs satisfying the promise, outputs a solution to the Unique Satisfaction Problem. On inputs which do not satisfy the promise, no requirements are made on the output. However, a crucial point is that, as devices in free theories are closed under composition, the output of the algorithm will always result in sensible probabilities regardless of the input. One is therefore free (no pun intended) to amplify the acceptance probability by running the algorithm many times and taking a majority answer. Hence, by Valiant-Vazirani it follows that $\mathbf{NP} \subseteq \mathbf{BGP}$, which using theorem 2.2 from chapter 2 gives $\mathbf{NP} \subseteq \mathbf{AWPP}$. □

See section 3.5 of the current chapter for a discussion on why the Valiant-Vazirani theorem alone is not sufficient to imply $\mathbf{NP} \subseteq \mathbf{AWPP}$.

3.3 Achieving the upper bound with a non-free theory

The main result of this chapter is—despite theorem 3.1—the construction of a non-free theory, satisfying causality, finite tomography, and tomographic locality, that has *exactly* the computational power of the class **AWPP**. See section 3.5 of the current chapter for a discussion on why the following theorem does not imply $\mathbf{NP} \subseteq \mathbf{AWPP}$.

Theorem 3.2. *There exists a non-free theory* **G***, satisfying causality, finite tomography and tomographic locality, such that*

$$\mathbf{BGP} = \mathbf{AWPP}.$$

The class **AWPP** contains problems for which an efficient quantum solution is unknown. Notable among these is the Graph Isomorphism problem, which asks for an efficient procedure to determine if two given graphs are equivalent. It is unknown whether a quantum computer can solve the Graph Isomorphism problem, and so our result provides a theory which can act as a “foil” to deepen our understanding of the limitations of quantum computers. Moreover, as discussed in the previous section, the promise version of **AWPP** contains the Unique Satisfaction Problem, which asks if a given Boolean formula has either a single satisfying assignment, or no satisfying assignment at all—promised one of these two cases is true. This is a very important problem and is closely related to many **NP**-complete problems [47].

We now provide an intuitive sketch of this construction, but defer the formal definitions and proofs to section 3.4. The proof starts by introducing a quasi-probabilistic model of computation, taking the form a Turing Machine with quasi-probabilistic transition weights with the constraint that the total weight of transitions from a given state must sum to +1. We refer to this model as an *Affine Turing Machine*. We show that the class of problems which can be efficiently solved with bounded error in this model perfectly captures the class **AWPP**. As illustrated schematically in Fig. 3.1, we then construct uniform poly-size circuits, in which the gates are certain affine transformations, that can simulate—and be simulated by—this affine Turing Machine, and hence **AWPP**. This construction results in a collection of closed circuits which correspond to the probability that the final result of the affine Turing Machine is “accept” or “reject” on inputs of different lengths. In the section 3.4 we prove that these closed circuits correspond to closed circuits in a causal, tomographically local (and tomographically finite) non-free theory. Thus completing the proof of theorem 3.2.

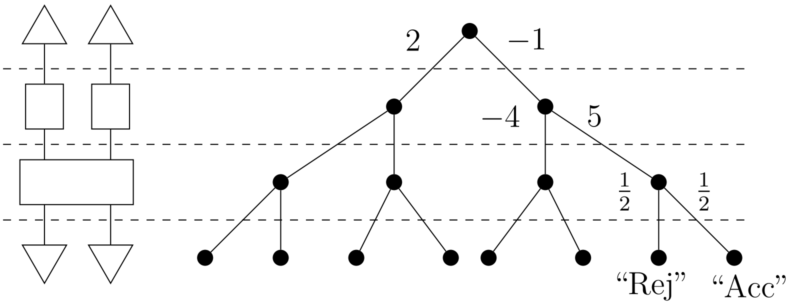


Fig. 3.1: Simulation of an affine Turing machine by an affine circuit

3.4 Proof of theorem 3.2

3.4.1 Affine Turing Machines

We define an *Affine Turing Machine* (AffTM) to be a non-deterministic Turing Machine, in which every transition has an associated real-valued (possibly negative) *weight*. We require that for each symbol being read, the total weight of transitions from a given (non-halting) state is $+1$. We also require that transition weights are rational.

We interpret AffTMs as a model of quasi-probabilistic computation, as follows. Given an AffTM $\mathbf{M}$ which halts in all paths in a finite number of steps, the *acceptance weight* $\alpha_{\mathbf{M}}(x)$ of that AffTM on an input x is the total weight of the accepting paths on input x. An AffTM $\mathbf{M}$ is *proper* if $0 \le \alpha_{\mathbf{M}}(x) \le 1$ for all inputs, and that it *decides a language $\mathcal{L}$ with bounded error* if furthermore $\frac{2}{3} \le \alpha_{\mathbf{M}}(x) \le 1$ for $x \in \mathcal{L}$, and $0 \le \alpha_{\mathbf{M}}(x) \le \frac{1}{3}$ for $x \notin \mathcal{L}$.

An AffTM is *efficient* if the number of computational steps it takes in any computational path on any input x is bounded by some polynomial in $|x|$. The first step towards Theorem 3.2 is to establish the following:

Lemma 3.3. *The class of languages decided by some efficient AffTM is equal to* **AWPP**.

The proof of this result is contained in following two sections.

3.4.2 Solving AWPP problems with an affine Turing machine

For such a language $\mathcal{L} \subseteq \{0,1\}^*$, let g be a **GapP** function[3] satisfying the definition of **AWPP** (i.e. definition 2.3 from chapter 2).

Also let $\mathbf{N}$ be the non-deterministic Turing machine whose accepting/rejecting branches determine the **GapP** function g, and T be the polynomial bounding the number of computational steps of $\mathbf{N}$ on its input. By standard results [55], we may require that $\mathbf{N}$ have the same number of non-deterministic transitions at each step, which we denote by $N \geq 1$, and that all computational branches of $\mathbf{N}$ have the same length on input x. We suppose that each transition of $\mathbf{N}$ is associated with some label $\ell \in \{1, 2, \ldots, N\}$: the computational branches of $\mathbf{N}$ are then in one-to-one correspondence with sequences $\{1, 2, \ldots, N\}^{T(|x|)}$. We may then consider an AffTM $\mathbf{M}$ which simulates $\mathbf{N}$, in the following sense:

1. $\mathbf{M}$ first makes a sequence of $T(|x|)$ non-deterministic transitions in which it writes symbols $\beta_1, \beta_2, \ldots, \beta_{T(|x|)} \in \{0, 1, 2, \ldots, N\}$ on the tape, constructing a string $\beta \in \{0, 1, 2, \ldots, N\}^{T(|x|)}$. The weights of these transitions are $+1$ for each choice $\beta_t \neq 0$, and $(1 - N)$ for each choice $\beta_t = 0$, so that the transition weights sum to $+1$.

2. In branches with one or more symbols $\beta_t = 0$, $\mathbf{M}$ transitions deterministically with weight $+1$ to a state REJECT. All other branches of $\mathbf{M}$ have weight $+1$ and record a string $\beta \in \{1, 2, \ldots, N\}^{T(|x|)}$ indexing some computational branch of $\mathbf{N}$. In these branches, $\mathbf{M}$ simulates the computational branch of $\mathbf{N}$ whose transitions are indexed by β.

3. For any branch in which the simulation of $\mathbf{N}$ rejects, $\mathbf{M}$ makes a non-deterministic transition to a state DAMPEN with weight -1, and to the REJECT state with weight $+2$. For the branches in which the simulation of $\mathbf{N}$ accepts, $\mathbf{M}$ transitions deterministically to DAMPEN with weight $+1$.

4. From the state DAMPEN, $\mathbf{M}$ makes a sequence of $p(|x|)$ non-deterministic transitions with weight $\frac{1}{2}$, in which it writes bits $\delta_1, \delta_2, \ldots, \delta_{p(|x|)}$ on the tape, thereby sampling a string $\delta \in \{0,1\}^{p(|x|)}$ uniformly at random. If $\delta = 11 \cdots 1$, $\mathbf{M}$ transitions to an ACCEPT state; in all other branches it transitions to the REJECT state.

[3] Recall from chapter 2 that a **GapP** function is a function $g : \{0,1\}^* \to \mathbb{Z}$ which computes the difference between the number of accepting and rejecting branches of some non-deterministic Turing machine $\mathbf{N}$

By the construction of the branch weights, $\mathbf{M}$ is an AffTM; and as the number of transitions that $\mathbf{M}$ makes is $O(T+p)$, it is efficient. By construction, the total weight of the branches which transition to the DAMPEN state is $g(x)$; sampling the string $\delta \in \{0,1\}^{p(|x|)}$ and rejecting unless $\delta = 11\cdots1$ ensures that the acceptance weight is $\alpha_{\mathbf{M}}(x) = g(x)/2^{p(|x|)}$. By hypothesis, this is bounded between 0 and 1, is at least $\frac{2}{3}$ if $x \in L$, and is at most $\frac{1}{3}$ otherwise. Thus $\mathbf{M}$ decides L with bounded error.

3.4.3 Simulating an Affine Turing Machine in AWPP

Suppose that $\mathbf{M}$ is a proper and efficient AffTM which has transitions with rational weights. Let N be the common denominator of the transition weights of $\mathbb{Q}$, $T \in O(\mathrm{poly}(n))$ the running time of $\mathbf{M}$ on an input of length n, and $m > 0$ be an integer such that $2^m \geq N^T$ and $2^m \geq \bigl(|u|N\bigr)^T$ for all transition weights u of $\mathbf{M}$; it follows that $m \in O(T)$. We may obtain an **AWPP** algorithm to approximately simulate $\mathbf{M}$, as follows. We define a non-deterministic machine $\mathbf{N}$, which simulates $\mathbf{M}$ in the following sense.

1. The machine $\mathbf{N}$ reserves some space on the tape to represent some weight $\Omega \in \mathbb{Q}$ for each branch. We call this the *recorded weight* of the branch.

2. Consider a transition made by $\mathbf{M}$, with weight $u = U/N$. To simulate this transition, the machine $\mathbf{N}$ replaces the recorded weight Ω with $\Omega' := U\Omega$, and then then simulates the actions (writing of symbols and movement of the tape head) performed by $\mathbf{M}$ in the original transition.

3. Once $\mathbf{N}$ has simulated the final transition of $\mathbf{M}$, it non-deterministically samples bits $a, b, c_0, c_1, \ldots, c_{m-1} \in \{0,1\}$. If $a = 1$, we negate Ω if and only if the simulated branch is one in which $\mathbf{A}$ rejects.

4. $\mathbf{N}$ determines whether to accept or reject, treating $c_{m-1}c_{m-2}\cdots c_1c_0$ as the binary expansion of an integer $0 \leq C < 2^m$, as follows.

   - If $C \geq |\Omega|$, we reject if $b = 0$, and accept if $b = 1$.
   - If $0 \leq C < |\Omega|$, we reject if $\Omega < 0$, and accept if $\Omega > 0$.

Consider the **GapP** function $f(x)$ of the machine $\mathbf{N}$. From Step 4, it is clear that if $C \geq |\Omega|$ in any particular branch, $\mathbf{N}$ accepts and rejects with equal measure, contributing nothing to $f(x)$. The significance of the contribution of any simulated branch of $\mathbf{M}$ is then in proportion to its recorded weight in $\mathbf{N}$, which in absolute value is $2N^T$ times its weight in $\mathbf{M}$ (arising from the systemic failure to divide the recorded weight by N at each of the T transitions, and from the two values of b). Let $\alpha_+(x)$ be the total weight of those accepting branches of $\mathbf{M}$ with positive weight, $\alpha_-(x)$ be the total (absolute value of) the weight of accepting branches with negative weight; and similarly for $\rho_+(x)$ and $\rho_-(x)$ for rejecting branches of positive and negative weight. Then $\alpha(x) := \alpha_+(x) - \alpha_-(x)$ is the acceptance weight and $\rho(x) := \rho_+(x) - \rho_-(x)$ is the rejection weight of $\mathbf{M}$ on input x. We decompose $f(x) = f_0(x) + f_1(x)$, where $f_0(x)$ is the contribution to the gap from branches in which $a = 0$, and $f_1(x)$ is the contribution to the gap from branches in which $a = 1$. We then have

$$f_0(x) = 2N^T\Big[\alpha_+(x) + \rho_+(x) - \alpha_-(x) - \rho_-(x)\Big] = 2N^T,$$

as $\alpha(x) + \rho(x) = 1$. In the branches where $a = 1$, the sign of the contribution from rejecting branches is negated, so that

$$\begin{aligned} f_1(x) &= 2N^T\Big[\alpha_+(x) - \rho_+(x) - \alpha_-(x) + \rho_-(x)\Big] \\ &= 2N^T\Big[2\alpha(x) - 1\Big], \end{aligned}$$

again using $\alpha(x)+\rho(x) = 1$. Thus, we have $f(x) = 4N^T\alpha(x)$. Let $g(x) = 4N^T$: as $\mathbf{M}$ is proper, then either $0 \leq f(x)/g(x) \leq \frac{1}{3}$ or $\frac{2}{3} \leq f(x)/g(x) \leq 1$. Thus, if $\mathbf{M}$ decides any problem with bounded error $\mathbf{N}$ is an **AWPP** algorithm for the same problem.

3.4.4 Constructing affine circuits

The next step towards Theorem 3.2 is to construct a family of circuits that can simulate an arbitrary AffTM. The final step will then be to show that the collection of all such circuit families is available in a specific non-free theory that satisfies tomographic locality and causality.

The construction of the circuits is based on that used by Yao in [79] to construct quantum circuits that simulate a quantum Turing Machine (and also on that of [91, 92] for circuits that simulate a probabilistic Turing machine). Let $\mathbf{M}$, as before, be a proper and efficient AffTM with alphabet Σ, set of states Q and transition amplitudes $\delta(q, a, \tau, q', a') \in \mathbb{Q}$ with $\tau \in \{\leftarrow, \circ, \rightarrow\}$; the symbols $\leftarrow$, $\rightarrow$ and $\circ$ are interpreted as the tape head of the AffTM moving to the left, moving to the right, and remaining stationary. Here δ is the transition weight of $\mathbf{M}$ to change to state q', print a' on the tape and move according to τ, if the machine is currently in state q and reading a. The condition on the weights in order for $\mathbf{M}$ to be an AffTM is: $\sum_{\tau,q',a'} \delta(q, a, \tau, q', a') = 1$ for all $q \in Q$, $a \in \Sigma$.

Denote the configuration of the AffTM at a specific step along one branch of the computation by the real vector

$$|s_0, q_0, a_0, \cdots, s_i, q_i, a, \cdots, s_{2t}, q_{2t}, a_{2t}\rangle,$$

where the index i denotes the ith cell of the tape and t is the run time of the AffTM (there are $2t+1$ cells, numbered from 0 to $2t$, instead of from $-t$ to t). Here s_i takes on value 0 when the head is not at cell i, value 1 when it is at cell i and the transition step has not occurred and value 2 when the head has just moved according to a transition and is now at cell i. Note that we can represent s_i with two bits. The label q_i denotes the internal state of the machine at cell i, so $q_i \in Q \bigcup \{\emptyset\}$ and it should be noted that $q_i = \emptyset$ if and only if $s_i = 0$. The label a_i denotes the alphabet character printed on cell i, i.e. $a_i \in \Sigma$. It is clear that l bits, where $l = 2 + \lceil \log(|Q| + 1) \rceil + \lceil \log(|\Sigma|) \rceil$, are required to represent the information at each cell. One can thus think of these vectors as being in one-to-one correspondence with elements in $\{0, 1\}^{(2t+1)l}$.

The transitions made along any one branch are represented by a sequence of these vectors. As the configuration of the AffTM at any stage of the computation is an affine combination of vectors of this form, the full transition step of the AffTM corresponds to an affine combination of such sequences.

We will now construct a uniform family of affine circuits that simulate this AffTM. Here, an affine circuit refers to an acyclic circuit consisting of a polynomial number of gates representing affine transformations acting on real vectors. We demand that (once a basis for the real vector space has been fixed) the matrices corresponding to these affine transformations have entries that can be computed efficiently, i.e. in poly-time, by an ordinary Turing Machine. We also demand that the description of the circuit can be computed in efficiently.

A specific affine circuit in this family will correspond to the concatenation of t identical sub-circuits, which we denote by B. Each sub-circuit B performs one step of the simulation. To construct these circuits, each tape cell of the AffTM is associated with a number of wires in a circuit. There are l wires, each associated with one bit, where $l = 2 + \lceil \log(|Q| + 1) \rceil + \lceil \log(|\Sigma|) \rceil$: the first two wires corresponds to the label s_i (recall s_i is represented by two bits); the next $\lceil \log(|Q| + 1) \rceil$ wires encode the classical bit-string describing an element of $Q \bigcup \emptyset$; the final $\lceil \log(|\Sigma|) \rceil$ wires encode the classical bit-string description of an element of Σ.

The construction of B will now be described. The basic building block of B is a gate G with $3l$ input wires and $3l$ output wires. B is built by cascading $2t - 1$ units of G, each sifting right by l wires, and at the end, adding a circuit I whose purpose

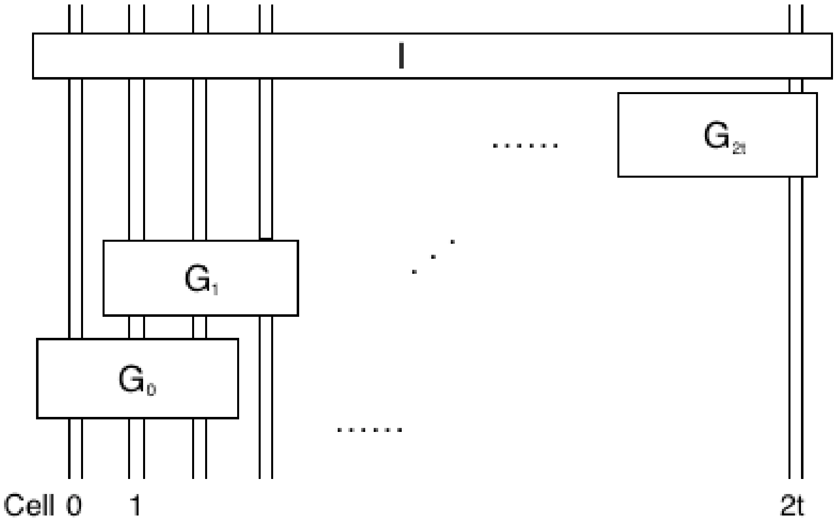


Fig. 3.2: Depiction of sub-circuit B.

is to change all s_i with value 2 to 1 and value 1 to 2. It is clear that I is an affine transformation and can be built using $O(t)$ gates whose function is to implement the change in s_i for a specific i. We denote the ith instance of G as G_i. See Fig. 3.2 for a pictorial representation of B.

The intuitive idea behind this construction is as follows. The $3l$ inputs to G should be thought of as describing the contents of three consecutive cells of the AffTM, including the information about the position of the head. We want G to transform the contents of these cells if the head is at the middle cell and the transition step has not occurred (i.e. $s_i = 1$ with i being the middle cell) according to how the AffTM would transform the contents. Thus we design G to act as follows:

1. For all $v = |s_{i-1}, q_{i-1}, a_{i-1}, s_i, q_i, a_i, s_{i+1}, q_{i+1}, a_{i+1})$ with $s_i \neq 1$, we have $G(v) = v$,

2. For $v' = |0, \emptyset, a_{i-1}, 1, q_i, a_i, 0, \emptyset, a_{i+1})$ we have

$$\begin{aligned} G(v') = &\sum_{q',a'} \delta(q_i, a_i, \leftarrow, q', a')|2, q', a_{i-1}, 0, \emptyset, a', 0, \emptyset, a_{i+1}) \\ &+ \sum_{q',a'} \delta(q_i, a_i, \circ, q', a')|0, \emptyset, a_{i-1}, 2, q', a', 0, \emptyset, a_{i+1}) \\ &+ \sum_{q',a'} \delta(q_i, a_i, \rightarrow, q', a')|0, \emptyset, a_{i-1}, 0, \emptyset, a', 2, q', a_{i+1}). \end{aligned}$$

We can think of G as a controlled affine transformation that does does nothing if the input has $s_i \neq 1$ and performs the transition step of the AffTM otherwise. Note that some linear combination of vectors with $s_i \neq 1$ can lead to the same output as when G is applied to a vector with $s_i = 1$. Thus in general G is not reversible, but this is not a problem as affine transformations are not reversible in general.

Note that as the configuration of the AffTM is an affine combination of vectors in one-to-one correspondence with elements in $\{0,1\}^{(2t+1)l}$, and, as we have defined the action of G (when tensored with the identity on cells on which it dos not act) on all such vectors, extending linearly uniquely defines G's action on all configurations of the AffTM.

Informally, B functions as follows: the cascading G's can be thought of as scanning over the contents of the AffTM's tape until the current position of the head is found and then implementing the transition step of the simulated machine. The circuit I then flips the value of s_i in so that the next simulation step can be performed. Thus B simulates one step of the AffTM and so, as the run time of the simulated machine is t, by concatenating t instances of B, the affine circuit can simulate each step of the AffTM. Recall that the configuration of a proper AffTM machine at step t is such that the sum of the weights of the accepting paths is positive. This ensures that the sum of the weights of the rejecting paths is also positive, as the accepting and rejecting weights must sum to one. Thus, the output of the circuit is a convex combination of the real vectors corresponding to the accepting and rejecting configuration. The probability to accept is then just the factor in front of the real vector corresponding to the accepting configuration.

In the next section, a slight variation on the above simulation will prove useful. The new simulation algorithm is the same as original discussed above except that, when the internal state of an AffTM enters into an accepting or rejecting state, the head of the AffTM will move to the first cell on the tape before halting. This ensures that in the convex combination of vectors output from the circuit, the state in the first cell will tell us whether the computation accepts or rejects, i.e. $q_0 = \text{Acc.}$ or $q_0 = \text{Rej.}$

3.4.5 A tomographically local theory

The construction in the preceding section resulted in a collection of closed circuits which correspond to the probability that the AffTM accepts or rejects given a specific input. Based on this, can we view these closed circuits as closed circuits in some causal and tomographically local non-free theory? As discussed in section 3.1, the correspondence between probabilities and closed circuits in a physical theory gives rise to a real vector space structure for the states, effects, and transformations. One must ensure that these emergent vector spaces have the structure one would expect given that the states, effects and transformations in the theory should correspond to the real vectors and matrices involved in the construction of the affine circuits. Both structures must match if we want to ensure the uniformity condition of the affine circuits and the fact that gates compose in parallel via the vector space tensor product, carries over to general theory. Moreover, these two structures must match in order for full tomography of the states and transformations to be possible within the theory.

For example, the vector

$$|s_0, q_0, a_0, s_1, q_1, a_1, \cdots, s_{2t}, q_{2t}, a_{2t})$$

lives in a $2^{(2t+1)l}$ dimensional vector space, but, from the point of view of the physical theory, the state corresponding to this vector lives in a two dimensional vector space; there are only two state preparations, those that lead to an accepting or rejecting output. In order to ensure the vector space dimensions match correctly, we introduce 'noisy' measurements consisting of 'noisy' effects that can be applied to single systems and which wash out any possible negative or super-normalised weights arising in the measurement process. First, we must introduce types for each system.

Recall that, at any specific step along one branch of the computation, the contents of cell i of the AffTM tape is described by a bit string $c_i \in \{0,1\}^l$, and these strings are in one-to-one correspondence with real vectors $|c_i)$. In order to consider these vectors as states in our theory, they must be assigned system types. We will denote the system associated to the vector $|c_i)$ as T_{A_i}, where T counts the number of steps the AffTM has undergone. Thus the system type associated with the state describing the contents of the ith cell after T steps in one specific branch is $|c_i)_{T_{A_i}}$. We have that $0 \leq T \leq t$, for t being the run time of the AffTM. We can think of the label T as specifying the 'vertical' position of the tree representing the AffTM computation and A_i as specifying the position along the tape.

Consider the following state of system 1_{A_i}, $|c_i)_{1_{A_i}}$, that is the state associated with the input to the AffTM computation. We demand that effects dual to these states satisfy ${}_{1_{A_i}}(d_i|c_i)_{1_{A_i}} = \delta_{c_i d_i}$, for $c_i, d_i \in \{0,1\}^l$. Given the construction of the real vectors from the section that constructed the affine circuits, the state corresponding to the entire contents of the tape at the very start of the AffTM computation corresponds to $|b)_{1_A} = \bigotimes_{i=1}^{2t+1} |c_i)_{1_{A_i}}$, for $b \in \{0,1\}^{(2t+1)l}$. Here, for simplicity, we are only considering an input of fixed size. The effects acting on these states are then ${}_{1_A}(e| = \bigotimes_{i=1}^{2t+1} {}_{1_{A_i}}(d_i|$, for d_i varying over bit strings from $\{0,1\}^l$. We define the set of states of system 1_A as the convex closure of the states described above and similarly for the effects.

To discuss systems of type T_A, for $1 < T < t$, we need to discuss the transformations defined by the gates G_i and I. We introduce types on the input and output ports of gate G and gate I (recall that I consists of $O(t)$ single system transformations) in such a way as to ensure the G's and I can only be composed to yield the circuit fragment B, as depicted in FIG. 3.2. There will be t versions of B, each with a different input and output type. For example, $B_{T_A}^{(T+1)_A}$ is a transformation that takes a system of type T_A to a system of type $(T+1)_A$.

States of system T_A are of the form

$$B_{T_A}^{T+1_A} \cdots B_{1_A}^{2_A} |b)_{1_A}$$

Now effects of the form ${}_{T_A}(e|$, for $e \in \{0,1\}^{(2t+1)l}$, cannot be allowed effects on system T_A as, due to the fact that the AffTM is allowed assign negative weights to certain transitions, ${}_{T_A}(e|B_{T_A}^{T+1_A} \cdots B_{1_A}^{2_A}|b)_{1_A}$ can in principle be negative. We now show how one can alter these vectors – by adding some 'noise' – in such a way as to ensure that the negative or super-normalised weights get washed out during the measurement process. Before we introduce noisy measurements, note that the following is an allowed effect on system T_A, which we call the unit effect and define by

$${}_{T_A}(U| = \sum_{e \in \{0,1\}^{(2t+1)l}} {}_{T_A}(e|.$$

Note that the expression

$${}_{T_A}(U|B_{T_A}^{T+1_A} \cdots B_{1_A}^{2_A}|b)_{1_A} = 1$$

follows from the fact that the amplitudes of a specific AffTM transition step sum to one.

We construct effects for the 'noisy' measurements, so that wherever they occur in the circuit, the measurement statistics do not reveal the negative coefficients suggested by our representation of the transformations. Assuming that we decompose the types into bit-strings, it suffices to define two effects $(a_0|_\nu = (0|D_\nu$ and $(a_1|_\nu = (1|D_\nu$, in terms of some stochastic 2×2 operator D_ν which provides a *veil of propriety*: that is, such that for any primitive gate G on k bits included in the theory, all of the coefficients of the matrix $(D_\nu^{\otimes k})G$ are in the interval $[0,1]$. Our strategy is to construct $D_\nu = \frac{1-p_\nu}{2}\left[\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right] + p_\nu \mathbb{I}_2$, where $\mathbb{I}_2$ is the 2×2 identity matrix, so that the effect of D_ν is to decrease the bias of any distribution on a single bit, and to determine how small p_ν must be to ensure that D_ν provides a veil of propriety. By 'bias' here we mean the difference between the two weights described by the inner products: $(i|x)$, for $i = 0, 1$, where $|x)$ is some state on system ν. If the bias is in the interval $[0, 1]$, then the weights form a probability distribution and we can consider $\{(a_0|, (a_1|\}$ to be a well-defined measurement acting on system ν. This motivates measuring the "maximum bias per bit" of each gate G, consisting of

$$\max_u \left|2\alpha - 1\right|^{1/k}$$

taken over all coefficients α of a k-bit gate G. An improper transition weight will be represented by biases-per-bit greater than 1.

The largest "improper" transition weight in the AffTM construction for **AWPP** algorithm is the transition weight $(1-N)$ associated with non-deterministically selecting a symbol $\beta_t = 0$ in a branching string, where N is the largest number of transitions of some NTM realising the **AWPP** algorithm. As the symbol $\beta_t \in \{0, 1, \ldots, N\}$ requires at least $\log_2(N + 2)$ bits to express, the bias-per-bit of these transitions are at most $(2N - 3)^{1/\log_2(N+2)} = 2^{\ln(2N-3)/\ln(N+2)}$. For $N > 1$, we may bound this from above by 3. The other improper weights in the construction are either -1 or $+2$, each having a maximum bias of 3 (or a maximum bias-per-bit of $3^{1/k}$ for k-bit gates with -1 or $+2$ as coefficients). We may then take 3 as an upper bound on the bias-per-bit of the gates in the affine circuit. To construct the noisy measurements, it suffices to choose $p_\nu = \frac{1}{3}$, which ensures that the maximum bias-per-bit of the gates $(D_\nu^{\otimes k})G$ as at most 1.

For system t_A the measurement corresponding to the AffTM accepting or rejecting its input, which consists of two effects, is allowed. To define this measurement we first add a subscript to the vectors corresponding to bit strings in $\{0, 1\}^{(2t+1)l}$, denoting the state of the AffTM machine in that configuration and we denote the subset of the set of states Q corresponding to the 'accepting' states by Q_{acc}. Consider the following two vectors:

1. ${}_{t_A}(e_{\text{acc}}| = \sum_{q \in Q_{\text{acc}}} {}_{t_A}(e_q|$, and;
2. ${}_{t_A}(e_{\text{rej}}| = \sum_{q \in Q/Q_{\text{acc}}} {}_{t_A}(e_q|$.

Recall that the configuration of the AffTM machine at step t is such that the sum of the amplitudes of the accepting paths is positive. This ensures that the sum of the amplitudes of the rejecting paths is also positive, as they must sum to one. Hence, the above two vectors define valid effects. Note also that ${}_{t_A}(e_{\text{acc}}| +_{t_A} (e_{\text{rej}}| =_{t_A} (U|$. This fact implies that the theory we are constructing is causal.

The conjunction of the accept/reject measurement provides a new measurement that can be performed on systems of type T_A, for $1 \leq T < t$. This new measurement corresponds to constructing the remainder of the circuit, by applying an appropriate number of gates G, and then performing the accept/reject measurement.

If we change the simulation algorithm of the AffTM slightly – as was briefly remarked upon at the end of the preceding section – we can ensure that the accept/reject measurement is not a joint measurement across all systems, but a single system measurement (composed with unit effects on all other systems). This implies that joint measurements across two identical copies of the same circuit are well-defined in our theory. The new simulation algorithm is the same as before except that when the internal state of an AffTM enters into an accepting or rejecting state, the head of the AffTM will move to the first cell on the tape before halting. This ensures we need only apply the accept/reject measurement to a single system t_{A_1}, in conjunction with unit effects on all other systems.

In summary:

1. System type 1_A:
   (a) **States:** All convex combinations of $|b)_{1_A} = \bigotimes_{i=1}^{2t+1} |c_i)_{1_{A_i}}$, i.e. $b \in \{0,1\}^{(2t+1)l}$.
   (b) **Effects:** Sums of effects ${}_{1_A}(e| = \bigotimes_{i=1}^{2t+1} {}_{T_{A_i}}(d_i|$, i.e. $e \in \{0,1\}^{(2t+1)l}$, with the property that ${}_{1_{A_i}}(d_i|c_i)_{1_{A_i}} = \delta_{c_i d_i}$, for all $i = 1, \ldots, 2t+1$, the effect corresponding to constructing the rest of the circuit and applying the accept/reject effect and the unit effect ${}_{1_A}(U| = \sum_{e\in\{0,1\}^{(2t+1)l}} {}_{1_A}(e|$.

2. System type T_A, for $1 < T < t$:
   (a) **States:** All convex combinations of $B_{T_A}^{T+1_A} \cdots B_{1_A}^{2_A}|b)_{1_A}$, for $1 < T < t$.
   (b) **Effects:** Unit effect ${}_{T_A}(U|$, 'noisy' effects and the effect corresponding to constructing the rest of the circuit and applying the accept/reject effect.

3. System type t_A:
   (a) **States:** All convex combinations of $B_{t-1_A}^{t_A} \cdots B_{1_A}^{2_A}|b)_{1_A}$
   (b) **Effects:** ${}_{t_A}(e_{\text{acc}}|$, ${}_{t_A}(e_{\text{rej}}|$, 'noisy' effects and the unit effect ${}_{t_A}(U|$.

The allowed closed circuits in the theory are those corresponding to the family of closed circuits which simulate **AWPP** and those formed by applying the local noisy measurements to the allowed individual states and gates in the theory. The latter set of closed circuits ensures one can perform full tomography of the states and transformations and thus ensure that each transformation in the theory has the same matrix representation as the gate from the affine circuit it corresponds to. This implies, among other things, that our theory satisfies tomographic locality as transformations in the affine circuits compose in parallel via the vector space tensor product. We have thus constructed a causal and tomographically local non-free theory (which satisfies finite tomography), completing the proof of theorem 3.2.

3.5 Discussion

One might wonder why the existence of a non-free theory satisfying **BGP** = **AWPP** does not immeadiately imply **NP** $\subseteq$ **AWPP**. Recall that the Valiant-Vazirani theorem [47] says that if one has both an efficient algorithm for solving the Unique Satisfaction Problem and the ability to probabilistically amplify polynomially-unlikely events, then one can solve any problem in **NP**. A crucial point in all of this is that the Unique Satisfaction Problem is a promise problem. That is, given the promise that the input is a Boolean formula with either a unique satisfying assignment or no satisfying assignment at all, then there exists an algorithm which outputs which of the two it was and satisfies the conditions of an **AWPP** computation on these inputs. When applying the Valiant-Vazirani procedure it may happen that an input to this algorithm does not satisfy such a promise. This was not an issue for free theories, as we saw in theorem 3.1. In a non-free theory however, it may not be the case that for every possible input—which may not satisfy the promise—the composite of the algorithm with that input is allowed in the theory. In the language of Affine Turing machines, if the input doesn't satisfy the promise, the output may not be proper, i.e. it may reveal the negative weights, and hence we cannot apply the Valiant-Vazirani theorem.

The distinction introduced in this chapter between free and non-free theories appears to be important for the study of computation in generalised probabilistic theories. The crucial distinction between free and non-free theories is that transformations in free theories are closed under composition, implying a bound on the set of states. This need not be the case in non-free theories. Could a quantum computer exploit this fact and efficiently simulate computation in all tomographically local free theories? If such a conjecture was borne out, it could shed light on which physical and structural features give rise to the quantum computational speed-up.

Recently, methods have been proposed that make use of quasi-probability distributions to classically estimate the output of a quantum computer [70]. These classical estimates converge on the true quantum output probabilities in a time quantified by the "negativity" of the quasi-probability distribution. The larger the negativity, the harder it is for a classical computer to estimate the output probability of a quantum computer. As we have provided an interpretation of the class **AWPP** in terms of quasi-probabilities, it would be interesting to determine if quantum algorithms can be constructed that estimate the output probability of this quasi-probabilistic computational model. In analogy with the classical estimation algorithms of [70] the quantum algorithms may converge to the true output probability at a rate governed by the negativity of the quasi-probability distribution. Determining how hard it is for a quantum computer to simulate **AWPP** would provide a way to determining if quantum theory is powerful for computation in the landscape of generalised probabilistic theories.

An interesting feature of the theory constructed in this chapter is that it satisfies the principle of causality. The main result of chapter 2 was that for any theory satisfying tomographic locality, whether or not causality is satisfied, efficiently solvable computational problems are contained in **AWPP**. Taken together, these results show that computational circuits in any non-causal theory can always be efficiently simulated by circuits in a causal theory. Hence, in the landscape of general theories, "acausality" does not appear to be a resource for computation.

Theorem 3.2 is reminiscent of a result encountered when quantum correlations are viewed in the context of the set of non-signalling theories [71]. This set consists of theories satisfying the *no-signalling* principle, ranked according to the strength of their correlations—quantified by the degree of violation of certain Bell Inequalities [71]. Quantum theory is ranked above classical theory, but Boxworld[4], which has the strongest correlations consistent with the no-signalling principle, is ranked above all other non-signalling theories. In the current chapter, we considered the set of theories satisfying tomographic locality and ranked them according to the power of their efficient computation. Classical theory is ranked below quantum theory (corresponding to the fact that quantum computers can efficiently simulate classical ones), but we showed there exists a theory with the strongest possible computational power and so is ranked above all other tomographically local theories, including quantum theory. In a sense, one can think of the theory constructed in this chapter as the analogue of a PR box for computation. This situation is illustrated schematically in Fig. 3.3

Moreover, Refs. [68, 69] have shown that methods employing quasi-probability distributions can simulate arbitrary non-signalling correlations. The quasi-probabilistic

[4] See chapter 1 for a discussion on Boxworld

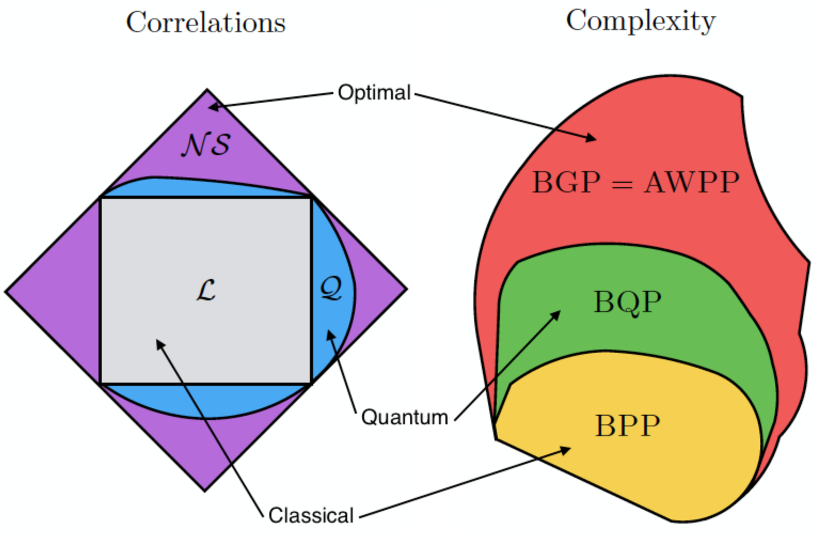


Fig. 3.3: The lefthand side schematically depicts the set of non-signalling theories, ranked in increasing order (from the center of the figure) of Bell inequlaity violation. The righthand side schematically depicts the set of all tomographically local theories, ranked in increasing order (from bottom to top) of the power of their efficient computation.

model of computation introduced here to construct a theory with maximal computational power bears an intriguing resemblance to these approaches, providing another similarity between the set of all non-signalling correlations and the computational landscape of general theories.

Interestingly, van Dam has shown [5] that communication complexity tasks in Boxworld can be solved trivially. Hence, quantum theory's computational and information processing capabilities appear to be sub-optimal in the broad landscape of generalised probabilistic theories. Studying situations that combine both computation and communication complexity in a non-trivial fashion may provide reasons why quantum theory is restricted in such a manner. In the next chapter, we use tools from the field of computational complexity, namely computation with advice and simple interactive proof systems, to investigate such situations.

Many attempts at providing reasonable physical principles that uniquely characterise the set of quantum correlations as a subset of the set of all non-signalling correlations have been made [77, 74, 76]. These principles, while not fully capturing the exact quantum boundary [75], have deepened our understanding of quantum correlations and provided connections between physical principles and information-theoretic advantages. Insights garnered from these connections have also lead to the development of Device-Independent Cryptography. So while investigating such connections has foundational interest, it has also been shown to have practical implications.

It seems prudent to ask the analogous question for the set of tomographically local theories: can the class of efficient quantum computation be characterised by some set of simple physical principles? That is, what is the minimal set of physical principles which perfectly captures the (top) boundary of the green region in Fig. 3.3. Such a characterisation would deepen our understanding of quantum computation and may also be of practical relevance; if one uncovers the necessary and sufficient physical requirements for universal quantum computation one could design algorithms that optimally take advantage of them. One approach to such a characterisation would be to find the minimal set of physical principles sufficient to derive the quadratic lower-bound to the search problem [44], which is saturated in the quantum case by Grover's algorithm. Any set of physical principles which derive it could be argued to capture some of the essence of quantum computation. This question will be the main focus of chapter 7.

Note again that the modified definition of generalised probabilistic theories presented in this chapter will not be used in subsequent chapters. In all other chapters—unless explicitly stated otherwise—the standard framework introduced in chapter 1 is used.

Chapter 4

Bounding the power of proofs and advice

In chapter 2, we showed that the class of problems a specific tomographically local generalised probabilistic theory can solve efficiently is contained in the classical complexity class **AWPP**. Moreover, in chapter 3, we explicitly constructed a non-free theory—satisfying both tomographic locality and causality—that can solve all problems in **AWPP** efficiently.

In section 3.5 of the previous chapter, it was noted that this situation is reminiscent of the one encountered when quantum correlations are viewed in the context of the set of non-signalling theories. Indeed, it was suggested that one could, in a sense, think of the theory constructed in the previous chapter as the analogue of a PR box for computation. The theory colloquially known as Boxworld, which exhibits PR box correlations and was discussed in chapter 1, was shown by van Dam [5] to be able to trivially solve any communication complexity task. This is due to the strength and complexity of the correlations present in Boxworld states. Quantum theory's information processing capabilities thus appear to be sub-optimal in the broad landscape of operational theories. Why then did Nature choose our universe to be quantum mechanical?

Taking a closer look at Boxworld reveals an intriguing partial answer to this question: reversible transformations in Boxworld are trivial—a fact first shown in [11] and strengthened in [95, 96]—and so any reversible computation in this theory can be easily simulated on a classical computer. Similarly, the non-free theory constructed in chapter 3 that achieves the maximal computation power appears to have sacrificed all other information processing capabilities. Indeed, the only non-trivial thing the theory seems to be abe to do is simulate **AWPP** computations[1]. These examples hint at the possible existence of a "trade-off", first suggested in [17], between the power of computation in a theory—that is, the richness of its dynamics—and its prowess in communication complexity tasks—that is, the amount of "useful" information that can be stored and extracted from its states. Could quantum theory be a "Goldilocks" theory that, in some sense, achieves the ideal trade-off in this regard? While it may not be the most powerful theory with respect to computation or communication complexity when considered separately, perhaps it has the optimal balance between powerful computation *and* powerful communication complexity.

We use the potential existence of such a trade-off as motivation to investigate situations in which both computation and communication complexity appear in a non-trivial manner in generalised probabilistic theories. In this chapter we investigate how simple physical principles bound the power of two different computational paradigms which combine computation and communication in a non-trivial fashion: computation with advice and interactive proof systems. Determining how computation and communication power vary as quantum theory is replaced by other operationally-defined theories may reveal some of the key physical features required for powerful computation and communication and is an interesting question independent of the motivation provided above.

4.0.1 Overview of the results

Computation with advice considers the situation where an efficient computer is supplemented with extra information, or *advice*, which, in classical computation, takes the form of a bit string and, in quantum computation, takes the form of a quantum state. The usefulness of this computational paradigm is that no uniformity[2] constraints are placed on the string or state embodying the advice—as is usually the case when one considers efficient computation—and one can thus attempt to encode solutions to hard problems in the advice. Aaronson was among[3] the first to study and set bounds on the power of quantum computation with (quantum) advice [102]. His primary motivation was a desire to investigate the question "How many classical bits can 'really' be encoded into n qubits?" from a complexity theoretic point of view.

[1] This is not to say that *any* theory with the computational ability of **AWPP** would suffer from the same condition

[2] As defined in chapter 2.

[3] Quantum computation with advice was first defined and studied in [101]

Aaronson noted [102] that quantum advice is quite closely related to quantum one-way communication[4], since one can think of an advice state as a one-way message sent to an algorithm by a benevolent "advisor". The class of decision problems which can be efficiently solved on a quantum computer with access to a quantum advice state is denoted **BQP/qpoly**, and Aaronson has shown [102] that $\mathbf{BQP/qpoly} \subseteq \mathbf{PP/poly}$. Based on the relation between quantum advice and quantum one-way communication, the size of the class **BQP/qpoly** can, in some sense, be thought of as a measure of prowess in communication tasks, or, intuitively speaking, as a measure of how much 'useful' information can be stored in a quantum state.

If the computational power of a general theory can be considered a measure of the richness of its dynamics, then the increase in computational power when supplemented with advice can be thought of—à la Aaronson in the previous discussion—as a measure of the information that can be stored in its states. In section 4.1 we provide rigorous definitions of the class of decision problems that can be solved in a specific generalised probabilistic theory when provided with a trusted advice state from that theory—which we call **BGP/gpoly** for a particular theory **G**. We show that in Boxworld, the class **BGP/gpoly** contains all decision problems and so is optimally powerful. Despite this, section 4.2 shows that theories with a certain amount of non-trivial reversible dynamics satisfy the same upper bound on the power of computation with advice as quantum theory. In particular, for theories **G** with non-trivial revsersible dynamics we show that $\mathbf{BGP/gpoly} \subseteq \mathbf{PP/poly}$. Boxword has no non-trivial reversible dynamics [11, 95, 96], and our result shows that when a theory *has* non-trivial reversible dynamics there is a limit on its prowess in certain communication tasks—as quantified by the size of the class **BGP/gpoly**. In a certain sense, one can view this result as a conceptual illumination of the conjecture concerning the existence of a trade-off between states and dynamics in physical theories, although it by no means establishes such a conjecture.

A key point in the above discussion is that one trusts the advice provider. That is, one trusts that the received advice contains the information the provider claims it does. In reality the provider could be malevolent and out to deceive the receiver.

[4] Quantum one-way communication can be described as follows: Alice has an n-bit string x, Bob has an m-bit string y, and together they wish to evaluate $f(x, y)$ where $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$ is a Boolean function. After examining her input $x = x_1 \dots x_n$, Alice can send a single quantum message ρ_x to Bob, whereupon Bob, after examining his input $y = y_1 \dots y_m$, can choose some basis in which to measure ρ_x. He must then output a claimed value for $f(x, y)$. We are interested in how long Alice's message needs to be, for Bob to succeed with high probability on any x, y pair.

If one cannot trust the provider, a computer must be used to check—or *verify*—that the provided advice is correct and this verification process requires non-trivial dynamics to implement. Thus, by learning how computational complexity changes as the amount of trust we have in the provider is varied, we enter into a regime where both prowess in communication tasks and computational power—corresponding to the existence of non-trivial dynamics—are simultaneously tested.

Within theoretical computer science, untrusted advice has been formally referred to as *proofs* and has a long history within computational complexity. For example, the famous class **NP** can be described as a proof system between an efficient, deterministic, classical computer, or verifier, and an all-powerful prover where the prover gives polynomially-sized proofs to the verifier. Here the verifier wishes to check if this proof is the correct solution to a particular problem. See appendix A for a rigorous definiton of the class **NP**. In quantum computing, a natural analogue of **NP** is the complexity class denoted **QMA**, for Quantum Merlin-Arthur. The question of what useful problems a quantum computer can solve when given a non-uniform quantum state as a proof from an untrusted source has led to surprising and beautiful connections between quantum computation and condensed matter physics [104].

In section 4.1, we give a rigorous definition of the class of problems for which a verifier with an efficient computer from a specific theory can solve when given proof states from that theory—which we call **GMA** for a particular theory **G**. We show, in section 4.3, that there exists a universal upper bound on **GMA** for all causal and tomographically local theories. In particular, we show that **GMA** $\subseteq$ **PP** for all **G** satisfying tomographic locality and causality. Note that Boxworld is an example of such a theory. Some results concerning the connection between trusted advice and proof verification in general theories are given in section 4.4.

4.1 Proofs and advice

In this section we provide generalisations of the definitions of classical (quantum) computing with advice and a type of classical (quantum) interactive proof system to the framework of general operational theories. For an overview of the classical and quantum definitions, see appendix B and C respectively. As with the definition of **BGP** in chapter 2, unless otherwise stated, the constants $(\frac{2}{3}, \frac{1}{3})$ can be chosen arbitrarily as long as they are bounded away from $\frac{1}{2}$ by some constant, or alternatively by some inverse polynomial in the size of the problem.

4.1.1 Definitions for general theories

Circuits from a uniform circuit family $\{C_x\}$ in some general theory are indexed by the string x that encodes the decision problem the theory is attempting to solve. In defining the class of efficient computation in a theory, the family $\{C_x\}$ is taken to consist of closed circuits from that theory. This will not be the case when advice and proofs are involved; in this paradigm, one is given both the problem instance x and a proof or advice state, so the constructed circuit C_x must have open system ports into which this state can be plugged. Henceforth we will assume that uniform circuit families consist of collections of circuits with a number of open input ports, which can grow as a polynomial in $|x|$, which we call the *auxiliary register.* Note that the choice of finite gate set determines the possible system types of the auxiliary register. Given this convention, we can define efficient computation with trusted advice in a specific general theory. Note in the following that we are assuming the causality principle.

Definition 4.1. *For a general theory* **G***, a language* $\mathcal{L} \subseteq \{0,1\}^n$ *is in the class* **BGP/gpoly** *if there exists a poly-sized uniform family of circuits* $\{C_x\}$ *in* **G***, a set of (possibly non-uniform) states* $\{\sigma_n\}_{n\geq 1}$ *on a composite system of size* $d(n)$ *for some polynomial* $d : \mathbb{N} \to \mathbb{N}$*, and an efficient acceptance criterion, such that for all strings*[5] $x \in \{0,1\}^n$*:*

1. *If* $x \in \mathcal{L}$ *then* C_x *accepts with probability at least* $2/3$ *given* σ_n *as input to the auxiliary register.*

2. *If* $x \notin \mathcal{L}$ *then* C_x *accepts with probability at most* $1/3$ *given* σ_n *as input to the auxiliary register.*

Here by "composite system of size $d(n)$", we mean that the number of systems, or open ports, of the auxiliary register—into which the advice state is input—increases as $d(n)$, for d a polynomial in the input size. Since, as mentioned in section 2.1.2 of chapter 2, there an efficient, deterministic, classical computer deciding acceptance, and each state σ_n has a classical pointer associated with it, classical advice can always be encoded into these pointers (of which there can be polynomially many in any poly-size circuit). Therefore, we can always give the lower bound **P/poly** $\subseteq$ **BGP/poly** $\subseteq$ **BGP/gpoly**, where the suffix **/poly** denotes classical advice.

Definition 4.2. *For a general theory* **G***, a language* $\mathcal{L} \subseteq \{0,1\}^n$ *is in the class* **GMA** *if there exists a poly-sized uniform family of circuits* $\{C_x\}$ *in* **G***, a polynomial* $d : \mathbb{N} \to \mathbb{N}$ *and an efficient acceptance criterion, such that for all strings* $x \in \{0,1\}^n$*:*

[5] i.e. strings x of length n, $|x| = n$.

1. *If $x \in \mathcal{L}$ then there exists a (possibly non-uniform) proof state σ on a composite system of size $d(n)$ such that C_x accepts with probability at least $2/3$ given σ as input to the auxiliary register.*

2. *If $x \notin \mathcal{L}$ then C_x accepts with probability at most $1/3$ given σ as input to the auxiliary system, for all states σ.*

As was stated at the start of this section, see appendix B and C for an overview of the classical and quantum definitions of computation with advice, and proof verification. Informally, for the specific case of quantum theory, the **G** in the nomenclature should be replaced with **Q** and **/gpoly** is replaced with **/qpoly**.

The existential quantifiers in the above definition of **GMA** rigorously capture the notion of a circuit having to "verify" the proof. Note also that advice states can only depend on the size of the input whereas proofs can, in general, be dependent on the inputs themselves. The amplification procedure of [107] that achieves exponential separation for the acceptance and rejection probabilities in **QMA**, at the expense of a polynomial increase in the size of the witness state, can be adapted in a straightforward fashion to provide a similar amplification procedure for **GMA**, for arbitrary **G**. Note that $\mathbf{BGP} \subseteq \mathbf{GMA}$ follows straightforwardly from the definitions. Also, via the same arguments given to lower bound the class **BGP/gpoly**, we can always give the lower bound $\mathbf{NP} \subseteq \mathbf{GMA}$.

It was proved in [107] that $\mathbf{QMA} \subseteq \mathbf{PP}$, and this was improved in [99] to $\mathbf{QMA} \subseteq \mathbf{A_0PP}$, (see also [106]). Aaronson and Drucker [103] have shown the following remarkable relation between these two classes:

$$\mathbf{BQP/qpoly} \subseteq \mathbf{QMA/poly}.$$

This says that one can always replace (poly-size) quantum advice by (poly-size) classical advice, together with a (poly-size) quantum proof[6]. Intuitively, this relation can be summed up as follows: one can always simulate an arbitrary quantum state ρ on all small circuits, using a different state $\widetilde{\rho}$ that is easy to recognize[7]. In section 4.3 we investigate whether this relation holds for general theories.

[6] Note that advice can encode solutions to even undecidable problems, any upper bound on an advice class will be another advice class.

[7] One can even take $\widetilde{\rho}$ to be the ground state of a local Hamiltonian [103].

4.1.2 Example: Boxworld

We now look at Boxworld with respect to our definitions of proofs and advice in general physical theories. Recall the definition of Boxworld from section 1.4.4 of chapter 1.

The 2-level PR box correlations discussed in chapter 1 can be extended to n-partite systems, where now for the jth party, $x_j \in \{0,1\}$ and $a_j \in \{0,1\}$ are the choice of measurement and its outcome respectively. There exists a state $|\rho_f)$ and effects $\{_j(x_j, a_j|\}$ for all j parties that produce the probabilities [108, 82]

$$(x_1, a_1|(x_2, a_2|...(x_n, a_n|\rho_f) = \begin{cases} \frac{1}{2^{n-1}}, \text{ if } \bigoplus_{j=1}^{n} a_j = f(x), \\ 0, \text{ otherwise} \end{cases}$$

where $\bigoplus$ represents summation modulo 2 and $f : \{0,1\}^n \to \{0,1\}$ is any Boolean function from the bit-string x with elements x_j. Therefore, if the state $|\rho_f)$ is prepared and local measurements described by effects $(x_j, a_j|$ made, a classical computer can compute the parity of all outcomes a_j and so we deterministically obtain the evaluation of Boolean function $f(x)$. This relatively straightforward observation gives us the following result.

Theorem 4.3. *There exists a generalised probabilistic theory* **G** *satisfying causality and tomographic locality, in which* **BGP/gpoly** = **ALL**, *where* **ALL** *is the class of all decision problems.*

Proof. Clearly **BGP/gpoly** ⊆ **ALL** is trivially true for Boxworld. The states $|\rho_f)$ can be used as advice states and, as all decision problems can be represented by Boolean functions, it follows that **ALL** ⊆ **BGP/gpoly**. □

Note as the above proof only requires the ability to prepare and measure states, it still goes through if we only restrict to reversible dynamics. If one considers the class **GMA** for Boxworld with only reversible transformations then we have **GMA** ⊆ **MA** since all reversible dynamics are trivial in this theory and can thus be simulated classically [11, 95, 96]. By trivial, we mean that the circuits in Boxworld only consist of making the local "fiducial" measurements $\{_j(x_j, a_j|\}$ on a state and performing classical post-processing on the outcomes. This process can be simulated by the prover giving the verifier the classical string of measurement outcomes similar to the approach of Lemma 2 in [97]. That is, while poly-size advice states in Boxworld can encode any Boolean function, the theory lacks the required reversible dynamics to efficiently verify this function is encoded in the state if the prover cannot be trusted.

4.2 Consequences of non-trivial dynamics for computation

In part 4.2.1 of this section, we show the existence of non-trivial dynamics implies that computation in that theory is at least as powerful as probabilistic classical computation: **BPP** $\subseteq$ **BGP**. Hence non-trivial dynamics imply non-trivial computational power. Furthermore in part 4.2.2, we show the existence of non-trivial dynamics implies a bound on the amount of "useful" information—quantified by the size of the class **BGP/gpoly**—that can be stored in general states.

4.2.1 Powerful computation from non-trivial dynamics

Definition 4.4. *A theory is said to be* non-classical *if, for at least one n-tuple of pure and perfectly distinguishable states* $\{|\sigma_i)\}_{i=1}^N$*, there exists a pure state* $|y)$ *such that* $(e_i|y) = p_i$ *for* $0 < p_i < 1$ *for all* i*, where* $\{(e_i|\}_{i=1}^N$ *is the measurement that distinguishes the* $\{|\sigma_i)\}_{i=1}^N$.

Before we present our result, we wish to emphasize that the result is highlighting the *intrinsic* computational power in a theory. As previously mentioned, in our framework we already have a classical computer that processes experimental data and, if a circuit in a theory **G** can produce random numbers, we can easily achieve the complexity class **BPP**. By talking about intrinsic computational power, we imagine reducing the power of our classical computer to perform extremely simple, non-universal classical computation. For example, the classical computer deciding the output of the computation could only output the (classical) pointer position of one of the measurement devices. Our result then shows that theories with a certain amount of non-trivial dynamics still decide any problem in **BPP**.

Theorem 4.5. *Let* **G** *be a causal, non-classical theory with at least two pure and distinguishable states that satisfies Permutability (as defined in section 1.3.4 of chapter 1.). Then*

$$\mathbf{BPP} \subseteq \mathbf{BGP}.$$

Proof. For **BPP** $\subseteq$ **BGP**, it is sufficient to show two things: that transformations of the general theory can simulate the action of any reversible Boolean function $f : \{0,1\}^n \to \{0,1\}^n$, and that it is possible to prepare a source of random bits. First, bit strings $x = x_1 \dots x_n$ can be represented by perfectly distinguishable pure states $|x) = |x_1) \otimes \cdots \otimes |x_n)$. Then, the first condition follows from Permutability:

since $\{|f(0\dots 0)), \dots, |f(1\dots 1))\}$ is a permutation of the tuple of pure and perfectly distinguishable states $\{|0\dots 0), \dots, |1\dots 1)\}$, there must exist a reversible transformation T_f such that $T_f|x) = |f(x))$.

For the second condition it suffices if there are circuits that can generate random bits. Consider the two pure and perfectly distinguishable states $|0)$ and $|1)$. Let $\{(e_0|, (e_1|\}$ be a measurement that distinguishes them, that is $(e_i|j) = \delta_{ij}$, for $i, j = 0, 1$. Non-classicality implies that there exists some pure state $|y) \notin \{|0), |1)\}$ such that $(e_0|y) = p$ and $(e_1|y) = 1 - p$, with $0 < p < 1$. Probabilities of $1/2$ can be generated by preparing two copies of $|y)$, implementing the measurement on each in parallel and assigning a value $y = 0$ or 1 to the outcomes 01 and 10 respectively[8].

□

4.2.2 Bounds on computation with advice in physical theories

Recall that a state is *mixed* if it is not pure and it is *completely mixed* if any other state refines it. That is, $|c)$ is completely mixed if for any other state $|\rho)$, there exists a non-zero probability p such that $p|\rho)$ refines $|c)$. Intuitively, one should be able to efficiently prepare a completely mixed state on a computer in any general theory. This follows because the completely mixed state can be prepared by performing any uniform state preparation and "forgetting" the outcome. Henceforth we shall assume that the completely mixed state—if it exists—is uniform.

Recall the definition of a bit-symmetry from section 1.3.4 in chapter 1. In any bit-symmetric theory with at least two pure and distinguishable states, it can be shown [49] that the group of reversible transformations acts *transitively* on the set of pure states. That is, given any two pure states $|\rho), |\sigma)$, there exists a reversible transformation T such that $T|\rho) = |\sigma)$. This fact can be used [15, 16] to prove the existence of a completely mixed state as the unique state—for a given system type—that is invariant under all reversible transformations.

Bit-symmetry is a powerful principle and has many useful consequences. Two more of which are:

1. Every bit-symmetric theory is *self-dual* [49]. That is, to every pure state $|\rho_e)$

[8] This argument is based on von Neumann's argument for turning two copies of a biased coin into one copy of an unbiased coin.

there is associated a unique pure effect $(e_\rho|$, and vice versa [9]. This association is achieved via an inner product $\langle\cdot,\cdot\rangle$, on the real vector space V generated by the set of states, as: $(e_\rho|\sigma) = \langle|\rho_e), |\sigma)\rangle$, for all states $|\sigma)$. Note that $\langle|\rho), |\rho)\rangle = 1$ for all pure states $|\rho)$.

2. Let $\||v)\|_{phy} = 2\max_{(e|} |(e|v)|$ and $\|v\|_E = \sqrt{\langle v, v\rangle}$, for v an arbitrary vector in V. Note that these are both norms. The norm $\||\rho) - |\sigma)\|_{phy}$ has a natural operational interpretation as the distinguishably of $|\rho)$ and $|\sigma)$. Bit-symmetry implies [114] that $\||\rho) - |\sigma)\|_{phy} \leq c\||\rho) - |\sigma)\|_E$, where $c = \||c)\|_E$ for $|c)$ the completely mixed state.

Using the above facts, we now prove a version of the "as good as new lemma" [10]—discussed in the quantum case in [102]—for all bit-symmetric theories. Before we state this lemma, we need to briefly introduce a notion of post-measurement state update rule for bit-symmetric theories. In this work applying a measurement to a state corresponds to a closed circuit—that is a probability. However, to discuss post-measurement states, this must be generalised slightly. A measurement will henceforth correspond to a laboratory device from some input state to the output post-measurement state, where the classical pointer denotes the outcome of the measurement. Consider the measurement $\{(i|\}$, consisting of pure effects $(i|$, and apply it to some state $|\rho)$. On observing outcome i, the state $|\rho)$ is updated to $|\rho_i)/(u|\rho_i)$ where $|\rho_i)$ is the unique pure state associated to $(i|$ via self-duality. This state update rule satisfies a natural *repeatability* condition: any state yielding outcome i with unit probability is left invariant by the update rule, thus repeated measurements always yield the same result. See [113] for more in-depth discussion of state update rules in general theories.

Lemma 4.6. *Given a two outcome measurement, consisting of the pure effects $\{(0|, (1|\}$, and a state $|\rho)$ such that $(0|\rho) = 1 - \epsilon$, for $\epsilon \geq 0$, the post-measurement state on observing outcome 0 satisfies*

$$\||\rho) - |\rho_0)\|_{phy} \leq c\sqrt{2\epsilon},$$

where $c = \||c)\|_E$ is the completely mixed state, in all bit-symmetric theories.

[9] The proof of this fact requires two further technical assumptions, both implicit in Ref. [49]. These are: the group of reversible transformations must be compact, and every mathematically allowed effect is physical.

[10] Also called the "gentle measurement lemma", which was independently proved by Winter in [117] and improved by Ogawa and Nagaoka in [118]

Proof. Recall in a bit-symmetric theory that $\||\rho) - |\sigma)\|_{phy} \leq c\||\rho) - |\sigma)\|_E$. We thus have

$$\begin{aligned}\||\rho) - |\rho_0)\|_{phy} &\leq c\||\rho) - |\rho_0)\|_E \\ &= c\sqrt{langle|\rho) - |\rho_0), |\rho) - |\rho_0)\rangle} \\ &\leq c\sqrt{2 - \langle|\rho), |\rho_0)\rangle - \langle|\rho_0), |\rho)\rangle} \\ &= c\sqrt{2 - 2\angle|\rho_0), |\rho)\rangle} \\ &= c\sqrt{2 - 2(0|\rho)} = c\sqrt{2\epsilon}.\end{aligned}$$

The first line follows from the definition of $\|.\|_E$, the second from the fact that $\||\sigma)\|_E \leq 1$ for all $|\sigma)$, the third from the symmetry of the inner product $\langle\cdot,\cdot\rangle$ and the last from the definition of self-duality. □

The above lemma states that if one outcome of a two-outcome measurement occurs with high probability on some state, then the post-measurement state after getting that outcome is "close" to the original state. We are now in a position to state the main result of this section. Before we do, let us fix the accepting criterion for computation with advice so that the acceptance function $a(\cdot)$ only depends on the measurement applied at the end of the computation. Moreover, we make the simplifying assumption that this final accept/reject measurement consists of pure effects.

Theorem 4.7. *Any causal, bit-symmetric, tomographically local theory* **G** *with at least two pure and distinguishable states satisfies*

$$\mathbf{BGP/gpoly} \subseteq \mathbf{PostBGP/poly} \subseteq \mathbf{PP/poly}.$$

The above theorem states that in theories with non-trivial reversible dynamics, there is a bound to how much useful information one an extract from any state. This result provides partial evidence for the existence of a trade-off between states and dynamics and can be seen as a natural converse to the results of [11, 95, 96]. Our proof is a slight variation of the original proof in the quantum case, due to Aaronson [102].

Proof. Begin by amplifying the success probability of **BGP/gpoly** on input x from $2/3$ to $1 - 1/2^{q(|x|)}$. This is achieved by running a polynomial number of copies of the circuit C_x in parallel and taking the majority answer, as described in chapter 2. Note that in this amplification scheme the total advice state is the (vector space) tensor product of advice states for each individual circuit. Recall that the completely mixed state $|c)$ is assumed to satisfy uniformity and that there exists a non-zero probability p such that $p|\sigma)$ is a refinement of $|c)$, for any $|\sigma)$. Uniformity implies that p can be well approximated by some rational $c/d^{w(|x|)}$, for c an integer and d a polynomial in the size of the input x (recall the proof of theorem 2.2 in section 2.5.2 of chapter 2).

Given any language $\mathcal{L} \in$ **BGP/gpoly** we now construct a **PostBGP/poly** algorithm that decides $\mathcal{L}$. Given some x, use the completely mixed state as the advice to the circuit C_x. Now, from the definition of **BGP/gpoly**, if $|c)$ cannot be used as advice to determine $x \in \mathcal{L}$, the circuit accepts with probability less than 1/3. Consider the post-measurement state $|c')$ of the auxiliary register after running C_x with advice $|c)$ *post-selecting* on the event that we succeeded in outputting the correct answer. If $|c')$ cannot be used as advice for all inputs, there exists some x' such that $C_{x'}$ succeeds with probability less than 1/3. As before, consider the post-measurement state of the auxiliary register after running $C_{x'}$ with advice $|c')$ post-selecting on outputting the correct answer. Continue in this fashion for some $t(|x|)$ stages, t a polynomial. Successful post-selection is guaranteed as the actual advice state refines $|c)$ with probability $c/d^{w(|x|)}$.

If, at any iteration of this process, we cannot find an x to move forward, we must be holding a state that works as advice for every input, and we can use it to run C_z on any input z, succeeding with high probability. Thus if the process halts after a polynomial number of iterations, we are done.

If the correct advice state $|\sigma)$ had been used in the computation, lemma (4.6) would imply the post-measurement state on observing the accepting outcome, $|\sigma_{acc})$, would—under the simplifying assumption that the accept/reject measurement consists of pure effects—satisfy:

$$\||\sigma) - |\sigma_{acc})\|_{phy} \leq c\sqrt{\frac{1}{2^{q(|x|)-1}}}.$$

As the completely mixed state $|c)$ is uniform, it follows that $c = \||c)\|_E \leq O(2^{m(|x|)})$ for m a polynomial. Therefore, $c/\sqrt{2^{q(|x|)-1}} = o(1)$. We thus have

$$\||\sigma) - |\sigma_{acc})\|_{phy} \leq o(1).$$

Therefore on each iteration of the above process, the correct answer is output with probability

$$\frac{c}{d^{w(|x|)}}\left(1 - o(1)\right).$$

This process has been designed so that the probability that $|c)$ can be re-used on each iteration and succeed at each stage is at most $1/3^{t(|x|)}$. Therefore, we have that

$$\frac{c}{d^{w(|x|)}}\left(1 - o(1)\right) \leq 1/3^{t(|x|)}.$$

Thus $t(|x|) \leq O(w(|x|))$ and we are done.

There thus exists a polynomial number of $x_1, \ldots, x_t$ such that, if $|a)$ is the post-measurement state after we start with $|c)$ and post-select on succeeding on each x_i in turn, $|a)$ is a good advice state for every string z. Provide the algorithm with this sequence of classical strings, along with the correct outcomes $b, \ldots, b_t$ for each of them. The algorithm then prepares $|c)$, uses it as advice and post-selects on getting outcomes $b, \ldots, b_t$. After this process we obtain the state $|a)$ and so all languages that can be decided in **BGP/gpoly** can also be decided in **PostBGP/poly** and thus, by tomographic locality and theorem 2.7 from chapter 2, in **PP/poly**. □

4.3 Bounds on the power of proofs in physical theories

In this section we will put a non-trivial bound on **GMA**. To state our result, we will again use the notion of a **GapP** function. We now define the class $\mathbf{A_0PP}$.

Definition 4.8. *A language $\mathcal{L}$ is in the class* $\mathbf{A_0PP}$ *if and only if there exists a* **GapP** *function f and an efficiently computable function T such that*

1. *for all $x \in \mathcal{L}$ $f(x) \geq T(x)$ and;*

2. *for all $x \notin \mathcal{L}$ we have $0 \leq f(x) \leq \frac{1}{2}T(x)$*

It known that the above class is contained in **PP** [99].

Fix the efficient acceptance condition for proof verification so that, in all uniform circuits, the acceptance function $a(\cdot)$ only depends on the measurement applied at the end of the computation to the non-auxiliary register. Moreover, we demand that the measurement applied to the auxiliary register consist only of unit effects. We make this choice to move closer to the standard quantum acceptance condition and to simplify the proof of the following theorem. We also make the simplifying assumption—routinely made in the literature—that all mathematically allowed states are physically allowed. That is, all vectors whose inner product with any effect is in $[0, 1]$ correspond to physical states.

Theorem 4.9. *For any generalised probabilistic theory* **G** *satisfying causality, tomographic locality and the assumption that all mathematically allowed states are physically allowed, we have that*

$$\mathbf{GMA} \subseteq \mathbf{A_0PP} \subseteq \mathbf{PP}.$$

Proof. Recall that any matrix M has a singular value decomposition given by $M = UDV^T$, where U, V are unitary (orthogonal if the matrix is real) matrices, V^T is the transpose of V and D is a diagonal matrix. The diagonal entries of D are all non-negative real numbers and are called the *singular values* of the matrix M. Note that the eigenvalues of the matrix $M^TM = VD^TDV^T$ are the squares of the singular values of M.

Let M_x be the matrix representation of the uniform circuit, including states and effects on the non-auxiliary register, on input x, $(u|$ be the (tensor product) of unit effects applied on the auxiliary register and $|\rho)$ be any arbitrary state (which can be non-uniform) input to the auxiliary register. Without loss of generality, one can pad this matrix (and row and column vector) with rows and columns of zeros to ensure it is square. The probability that the circuit accepts the string x is given by $(u|M_x|\rho)$. It will now be shown that this probability is upper bounded by the largest singular value of the matrix M_x. Consider the following

$$(u|M_x|\rho) = (u|UDV^T|\rho) \leq \sigma_{max}(u|UV^T|\rho),$$

where σ_{max} is the largest singular value of M_x. Now UV^T is a unitary matrix and so can be decomposed as follows $UV^T = WD'W^T$, where W is another unitary matrix and D' is a diagonal matrix consisting of the eigenvalues of UV^T, recall that these eigenvalues all have absolute value 1. Thus,

$$(u|M_x|\rho) \leq \sigma_{max}(u|WD'W^T|\rho) \leq \sigma_{max}(u|\rho) \leq \sigma_{max},$$

where the second inequality follows from that fact that the entries of D' have absolute value 1 and that W is unitary and the third inequality follows as $(u|\rho) \leq 1$.

Now as the squares of the singular values are the eigenvalues of the (positive definite) matrix $M_x^TM_x$, we have that

$$(\sigma^2_{max})^d \leq \mathrm{Tr}\big((M_x^TM_x)^d\big) \leq 2^n(\sigma^2_{max})^d,$$

where 2^n is the number of entries on the diagonal of $M_x^TM_x$, n is a polynomial in $|x|$ an d is an arbitrary natural number. Let d be a polynomial in $|x|$ that takes values in the natural numbers and assume without loss of generality that it grows faster than the polynomial n, we will need this requirement later.

The matrix M_x satisfies the uniformity condition, and it was shown in section 2.5.2 of chapter 2 that the entries of all such matrices are **GapP** functions. By the closure properties of **GapP** (again see chapter 2) functions the entries in the matrix $(M_x^TM_x)^d$

are also **GapP** functions. Using an argument similar to that in [99], $\mathrm{Tr}\big((M_x^T M_x)^d\big)$ can be straightforwardly shown to be a **GapP** function, denote it by $f(x)$. So, from the definition of **GMA**, we have that $f(x) \geq \sigma_{max}^{2d} \geq \left(\frac{2}{3}\right)^{2d}$ for all x in the language.

Now the vector that achieves the bound of σ_{max} is the right singular vector of M_x with singular value σ_{max}, which we denote by $|\sigma)$. If this vector is a physical state then we are done, as it follows from the definition of **GMA** and an argument similar to the one above that $f(x) \leq \frac{1}{2}\left(\frac{2}{3}\right)^{2d}$ for all x not in the language. If this vector is not a physical state then we have a bit more work to do.

Towards this end, consider the following. We are free to re-parametrise (e.g. see page 7 of [100]) the set of states by an affine transformation $\phi : \mathbb{R}^m \to \mathbb{R}^m$, where $\mathbb{R}^m$ is the (smallest) real vector space that contains the set of states, as follows:

$$|\rho) \to |\widetilde{\rho}) = \phi|\rho), \quad (a| \to (\widetilde{a}| = (a|\phi^{-1}$$
$$\text{and,} \quad M_x \to \widetilde{M_x} = \phi M_x \phi^{-1}$$

as this does not change the probabilities, i.e. $(a|M_x|\rho) = (\widetilde{a}|\widetilde{M_x}|\widetilde{\rho})$. Now, as an affine transformation can be thought of as a translation followed by a scaling, choose ϕ so that the Euclidean unit ball is contained in the re-parametrised state space (just translate the original state space and scale it appropriately to ensure this, noting that translations and scaling are reversible). As the singular vectors of every matrix are unit vectors, without loss of generality they are contained in this unit ball. Under the assumption that all mathematically allowed states are physically allowed ensures these singular vectors are physical states. Thus $\sigma_{max} = (\widetilde{u}|\widetilde{M_x}|\sigma)$, where $(\widetilde{u}|$ is the unique deterministic effect. The causality principle ensures that one can *re-normalise* any state, that one can scale each state $|\widetilde{s})$ so that $(\widetilde{u}|\widetilde{s}) = 1$ [15]. So for x not in the language we have $\sigma_{max} = (\widetilde{u}|\widetilde{M_x}|\sigma) \leq 1/3$.

It follows that

$$f(x) \leq 2^n \sigma_{max}^{2d} \leq 2^n \left(\frac{1}{3}\right)^{2d} \leq \frac{1}{2}\left(\frac{2}{3}\right)^{2d},$$

where the first inequality follows from $\mathrm{Tr}\big((M_x^T M_x)^d\big) \leq 2^n (\sigma_{max}^2)^d$ and the last inequality follows from the fact that, for d increasing sufficiently faster than n, we have $2^{n+1} \leq 4^d$.

Thus, for a language $\mathcal{L}$ in **GMA** we have

1. for all $x \in \mathcal{L}$ there exists a **GapP** function f such that $f(x) \geq \left(\frac{2}{3}\right)^{2d}$ and;
2. for all $x \notin \mathcal{L}$ we have $f(x) \leq \frac{1}{2}\left(\frac{2}{3}\right)^{2d}$,

and so we have that $\mathbf{GMA} \subseteq \mathbf{A_0PP}$.

□

4.4 Relating proofs and advice?

The following relation, discussed in section 4.1, $\mathbf{BQP/qpoly} \subseteq \mathbf{QMA/poly}$, captures an intriguing feature of proofs and advice in quantum theory: one can always replace quantum advice with classical advice together with a quantum proof. Here we study the relation

$$\mathbf{BGP/gpoly} \subseteq \mathbf{GMA/poly}, \tag{4.1}$$

in general theories. Note that the relation is satisfied in classical computation:

$$\mathbf{BPP/rpoly} = \mathbf{P/poly} \subseteq \mathbf{NP/poly} \subseteq \mathbf{MA/poly},$$

where $\mathbf{BPP/rpoly} = \mathbf{P/poly}$ was shown in [116, 105]. Clearly the relation in (4.1) is then not uniquely satisfied by quantum theory, but one could ask whether quantum theory is the most computationally powerful theory in which (4.1) is satisfied?

Using these observations as motivation we obtain the following corollary of theorem 4.9.

Corollary 4.10. *There exist general theories* $\mathbf{G}$ *satisfying tomographic locality and causality such that* $\mathbf{BGP/gpoly} \not\subseteq \mathbf{GMA/poly}$.

Proof. Firstly, we can use theorem 4.9 to conclude that $\mathbf{GMA/poly} \subseteq \mathbf{PP/poly}$ and by a counting argument $\mathbf{PP/poly}$ is strictly contained in $\mathbf{ALL}$. From theorem 4.3, there exists a theory $\mathbf{G}$ such that $\mathbf{ALL} = \mathbf{BGP/gpoly}$ and so we do not have $\mathbf{BGP/gpoly} \subseteq \mathbf{GMA/poly} \subseteq \mathbf{PP/poly}$ for this theory. □

Motivated by the above corollary we can say something non-trivial about theories where $\mathbf{BGP/gpoly} \not\subseteq \mathbf{GMA/poly}$. Consider the case of using a polynomially-sized circuit from a specific theory, built from any fixed gate set in that theory, to prepare an arbitrary, but polynomially large, state in the theory. Given this set-up, we can prove the following result.

Theorem 4.11. *In any general theory* $\mathbf{G}$ *with*

$$\mathbf{BGP/gpoly} \not\subseteq \mathbf{GMA/poly}$$

there exist states (of polynomial size) that cannot be prepared using an efficient circuit built from any gate set in the theory.

Proof. Assume toward contradiction that all states can be prepared using an efficient circuit built from any gate set in the theory. Thus, as there must exist a classical description of each circuit, any advice state from this theory can be replaced with the *classical* advice that specifies the description of the circuit that efficiently prepares the given advice state. We thus have

$$\mathbf{BGP/gpoly} \subseteq \mathbf{BGP/poly} \subseteq \mathbf{GMA/poly},$$

which is a contradiction. There must therefore exist at least one state that cannot be prepared efficiently in this theory. □

Thus in theories that do not satisfy

$$\mathbf{BGP/gpoly} \subseteq \mathbf{GMA/poly},$$

the dynamics are not rich enough to prepare the states that contain a large amount of "useful" information. This is not to say that in theories satisfying this relation every state can be efficiently prepared, it is just that in theories violating the relation this assertion can be proved *directly* from the violation. As a side remark, within the theorem proof we have proven that **BGP/poly** is *strictly* contained in **BGP/gpoly** for theories **G** where $\mathbf{BGP/gpoly} \nsubseteq \mathbf{GMA/poly}$. It is presently unknown if quantum advice is strictly stronger than classical advice for quantum computers.

In addition to proving $\mathbf{BGP/gpoly} \subseteq \mathbf{GMA/poly}$, Aaronson and Drucker proved what they called a "Quantum Karp-Lipton" theorem [103]. The Karp-Lipton theorem states that if $\mathbf{NP} \subseteq \mathbf{P/poly}$ then the polynomial hierarchy collapses to its second level, which is believed to be unlikely [109]. The Quantum Karp-Lipton theorem states that if $\mathbf{NP} \subseteq \mathbf{BQP/qpoly}$ then the second level of the polynomial hierarchy is contained in $\mathbf{QMA}^{\mathbf{PromiseQMA}}$ [11], which is also thought to be unlikely [103]. We refer the reader to the original works for further details but we only wish to highlight that, due to theorem 4.3, there exist theories **G** where $\mathbf{NP} \subseteq \mathbf{BGP/gpoly}$ is necessarily satisfied. Therefore, we cannot obtain a "Generalised Karp-Lipton" theorem where unlikely consequences are expected from assuming $\mathbf{NP} \subseteq \mathbf{BGP/gpoly}$.

4.4.1 Related work

Evidence for the existence of a general trade-off has also appeared in recent work which has considered theories satisfying the no-signalling condition from the point-of-view of interactive proofs. The Merlin-Arthur game is an example of an interactive

[11] Here **PromiseQMA** is the same as **QMA** except there is a "promise" on the inputs, i.e. all the inputs satisfy some property.

proof. Another example is a multi-interactive prover (**MIP**) system where more than one of these all-powerful provers sends classical bit-strings to a probabilistic classical computer verifier [110]. Just as in the Merlin-Arthur game, the provers cannot be trusted. However, these provers are not permitted to communicate with one another. A quantum generalisation of this is to allow the provers to share entangled quantum states. In work by Ito and Vidick [111], in this quantum generalisation of **MIP** it is possible for the verifier to efficiently compute problems in the class **NEXP** which is the class of problems evaluated by a non-deterministic computer running in time exponential in the size of the input. However, recent work by Kalai, Raz and Rothblum [112] has shown that if the provers share resources that satisfy only the no-signalling principle (such as Boxworld), then the problems that can be solved in such a model are actually contained in the class **EXP**. Since **EXP** $\subseteq$ **NEXP**, in a theory with states more non-local than quantum mechanics these interactive proof systems have *less* computational power, unless **EXP** = **NEXP**.

4.5 Discussion and conclusion

The results in this chapter provide another example where the best known upper bound on a quantum complexity class (in this case, **QMA**) follows from very minimal assumptions on what constitutes an operational theory. This raises the question of whether better bounds can be derived in the quantum case by exploiting some of the structure unique to quantum theory.

One can interpret the fact that **GMA** $\subseteq$ **PP** holds for tomographically local and causal **G** as partial evidence for the existence of a general trade-off between states and dynamics in operational theories. Indeed, if *both* the computational power and prowess in communication tasks of a theory **G** increase, then, intuitively speaking, so too may the size of the class **GMA**. If both increase by a large amount it is conceivable that the bound **GMA** $\subseteq$ **PP** may no longer hold. When viewed this way, this upper bound on **GMA** appears to suggest a limit to the power of computation in a theory given its prowess in communication tasks, and vice versa. Since, otherwise, one may no longer expect the upper bound to hold. This is just one possible interpretation of this result, indeed it could turn out to be the case that computation and communication power *are* in fact positively correlated and all our result implies is that **BGP** and **GMA** both cannot exceed **AWPP** and **PP** respectively. The status of the "trade-off conjecture" is still open, but the results of this chapter suggest that studying situations which involve both computation and communication complexity may suggest a path to its resolution.

While the definitions of advice and proof verification presented in this chapter can be applied to any theory in the framework, they seem to intuitively encode a notion of causality. Note that in a non-causal theory, circuits do not have any particular "direction" and so inputting a given state at the "start" of the computation is not the most natural situation one could consider. Instead of receiving an advice *state*, a more natural situation might be to receive an advice *circuit fragment*—consisting of either a state, transformation or measurement—which can be plugged into the circuit as it is being built. It would be interesting to determine if this more general definition coincides with the standard one in extensions of quantum theory with indefinite causal structure [61, 62].

On a final note, it would be fascinating if the analysis of computation in generalised probabilistic theories could say something concrete about quantum computing. In an analogous fashion, tools from quantum theory have been used to prove results in *classical* computer science, see [115] for a nice review of such results. We speculate that by understanding quantum theory better within the framework of more general theories we can use tools from the latter to prove results in the former.

Chapter 5

Higher-order interference

Over the course of the previous three chapters we have investigated the connections between physical principles and computation, using the language of complexity classes to derive general bounds on the power of computation in generalised probabilistic theories. However, as mentioned in the introduction to this thesis, much of quantum computing is concerned not so much with the high-level view offered by complexity classes, but instead with the construction of concrete algorithms to solve specific problems. Quantum interference between computational paths has been posited [84] as a key resource behind the computational "speed-ups" offered by many quantum algorithms, such as Grover's search algorithm [149]. However, as first noted by Sorkin [124, 125], there is a limit to this interference—at most pairs of paths can ever interact in a fundamental way. Sorkin has defined a hierarchy of possible interference behaviours where classical theory is at the first level of the hierarchy and quantum theory belongs to the second. Informally, the order in the hierarchy corresponds to the number of paths that have an irreducible interaction in a multi-slit experiment. Could more interference imply more computational power?

This conjecture will be investigated over the course of the remaining chapters of this thesis. The current chapter provides an overview of the literature on higher-order interference and investigates two proposed extensions of quantum theory from the perspective of their interference behaviour. The two theories in questions are the theory of Density Cubes proposed by Dakić, Paterek and Brukner, which has been shown to exhibit third-order interference in the three slit set-up, and the Quartic Quantum Theory of Życzkowski. This investigation clarifies the impact of these two generalised theories to ongoing experimental tests for higher-order interference and explores potential information-theoretic consequences of post-quantum interference in concrete theories. In particular it highlights an ambiguity in the current definition of higher-order interference, which shall be remedied in chapter 6. In chapter 6 we show how some of the essential components of computational algorithms arise from physical principles alone and use these tools to begin an investigation of the relationship between interference behaviour and computational power. Finally, in chapter 7 we use the results of chapter 6 to to investigate how Grover's speed-up depends on the order of interference in a theory.

5.1 Background and introduction

The predictions of quantum theory are the most accurately tested of any physical theory in the history of science. Nevertheless, it may turn out to be the case that quantum theory is only an effective description of a more fundamental theory whose predictions deviate from those of quantum theory in certain energy regimes or in sufficiently sensitive experimental set-ups. It is thus of the utmost importance that fundamental tests of the validity of quantum theory be performed. Such tests take a characteristically quantum prediction and probe the limits of its accuracy in different experimental situations. One such prediction, currently under experimental investigation [131, 132], is the limitation of quantum theory to second, as opposed to higher, order interference in n-slit experiments.

Higher-order interference was first described by Sorkin [124] who noted that quantum theory is limited to having only second-order interference. Informally, this means that the interference pattern obtained in a three—or more—slit experiment can be written in terms of the two and one slit interference patterns that are obtained by blocking some of the slits; no genuinely new features result from considering three slits instead of two. This is in stark contrast to the existence of second-order, i.e. quantum-like, interference, where the two-slit interference pattern *cannot* be written as a sum of the one slit patterns obtained by blocking each one of the slits. This was first made precise in the context of quantum measure theory [125], where moving from classical to quantum theory can be seen as a weakening of the Kolmogorov sum rule to allow for second (but not third, or higher) order effects.

Restriction to only second-order interference appears to be a characteristically quantum phenomena and many other 'quantum-like' features, which, at first glance, appear to be unrelated to interference, can be derived from it. For example: limiting correlations [127, 128] to the "almost quantum correlations" discussed in [75], and bounding contextuality [74]. Additionally a lack of third-order interference was also used by Barnum, Müller and Ududec [72] as a postulate in their reconstruction of quantum theory.

The natural question that arises from this discussion is *why* does quantum theory only exhibit second-order interference? It may strike some as odd that there is a limit to the non-classicality of quantum theory. Why is nature strange, but not excessively so? Does the existence of genuine third-order interference imply incredible computational power that we could consider physically unreasonable?

Barnum, Müller and Ududec provide an operational definition [72] (see also [78]) of higher-order interference that is—at least at first sight—applicable in any generalised probabilistic theory. Given this definition, one can attempt to construct a theory that exhibits higher-order interference in the hope of using it as a 'foil' to quantum theory. Such a foil theory would hopefully shed some light on possible pathological—or at least undesirable—features of higher-order interference and thus provide reasons 'why', in some sense, quantum theory should be limited to second-order interference. Currently, there are no 'complete' generalised probabilistic theories that exhibit third-order interference. There are particular state spaces [129] that have higher-order interference but these are of a fixed dimensionality and composition is not discussed, additionally they have a highly restricted set of dynamics when compared to quantum theory.

It would be of particular interest if there was a theory that exhibited higher-order interference and which contained quantum theory as a limiting case. Yet, if such a theory exists, there should be some mechanism by which the magnitude of effects unique to this theory are suppressed, thus explaining why quantum theory is such a good effective description of the world. This mechanism would be analogous to the process of decoherence, which induces the quantum-classical transition and which makes observation of genuine quantum effects difficult to experimentally detect. Hence the mechanism by which an extension of quantum theory reduces to standard quantum theory is called *hyper-decoherence*[1]. Any well-defined theory that extends quantum theory should provide a mechanism for hyper-decoherence. Experimental bounds have been found limiting the possible amount of third (or higher) order interference [131, 132] and these place stringent bounds on the hyper-decoherence time of potential extensions of quantum theory.

Ududec, Barnum and Emerson have shown [78] that the absence of third-order interference is equivalent to the ability to perform full tomography of any state using only measurements consisting of two-slit experiments, i.e by only performing measurements on two dimensional subsystems[2]. It follows that any theory which exhibits

[1] See [122] for a more in-depth discussion of hyper-decoherence.

[2] i.e. by only performing measurements of the form $a\langle i| + b\langle j|$.

genuine third-order interference, and aims to be an extension of quantum theory, requires more parameters to specify an n-level system than are required to specify an n-level quantum system. Intuitively then, one can think of the dimension of the subspace upon which one needs to perform measurements to do complete tomography as corresponding to the order of interference.

Guided by this, Dakić, Paterek and Brukner [121] have proposed a method to construct a theory which exhibits third-order interference and which extends standard quantum theory. In section 5.3.3, we show that this construction gives rise to a sensible notion of hyper-decoherence which leads to the emergence of quantum theory in particular cases—analogous to the emergence of classical physics from quantum theory. In section 5.3.4 we also show that this construction provides an advantage over quantum theory in a certain computational task, foreshadowing results in chapters 6 and 7. Despite all these nice features, in section 5.3.5 we demonstrate that this approach—as it is currently presented—does not lead to a well-defined physical theory by showing the axioms defining the state space are insufficient to uniquely characterise the theory. It is therefore suggested that one can view the theory of Density Cubes more as a framework for developing operational theories than a unique theory.

Another feature of tomography in the generalised probabilistic theory framework is discussed by Hardy in [18], where a hierarchy of theories are presented and shown to satisfy the relation $K = N^r$, where K is the number of effects whose statistics are required to completely determine a state, N is the dimension of the system, and r is a positive integer specifying the level in the hierarchy. The case $r = 1$ corresponds to classical theory and $r = 2$ to quantum theory[3]. For $r > 2$ one may expect—based on the results of [78] discussed above—that tomography on these higher dimensional subspaces leads to higher-order interference. The results of [78] suggest that the rth level of this hierarchy, i.e. $K = N^r$, should exhibit rth-order interference, but no higher.

Życzkowski has developed a theory [122] satisfying $K = N^4$, which extends quantum theory, and so provides a candidate for a theory of higher-order interference. In section 5.4 it will be shown that Życzkowski's $K = N^4$ theory does not suffer from the problems of Dakić *et al.'s* construction; there is a unique state space associated with the theory and all transformations are physical. Furthermore, this theory does indeed exhibit third—and higher—order interference. In fact, every n-level system in this theory exhibits nth-order interference, for all n. This is somewhat surprising

[3] Note that we are allowing sub-normalised states, hence quantum theory satisfies $K = N^2$ rather than $K = N^2 - 1$.

and unexpected as, based on the discussion in the previous paragraph, one would expect this theory to exhibit at most 4th-order interference. Another surprising, and somewhat worrying, feature of interference in this theory is the fact that the existence of higher-order interference stems from a somewhat artificial and operationally unmotivated choice. Blocking some subset of the slits corresponding to apertures in the physical barrier describing an n-slit experiment should uniquely define a measurement. Życzkowski's theory does not posses this feature: there exist (at least) two well-defined measurements that correspond to blocking the same subset of slits in an n-slit experiment. One of these measurements results in higher-order interference, the other does not.

Arguably, both of these features arise from a limitation of Barnum, Müller and Ududec's definition of higher order interference rather than a genuine phenomenon; there should be a unique measurement that corresponds to opening any subset of slits, and this does not appear to happen without further constraints on the theory.[4] Thus one should not consider Życzkowski's theory as an example of higher-order interference in the sense originally meant by Sorkin, but rather a demonstration of the challenges of applying his original definition to arbitrary generalised probabilistic theories. Thus to begin to understand the reason why, in some sense, quantum theory is limited to second-order interference, we first need a definition of higher-order interference that is applicable to, and makes good operational sense in, arbitrary generalised probabilistic theories. Such a definition is provided at the start of chapter 6

Finally, in section 5.4.3.1, we briefly comment on the type and strength of correlations allowed in Życzkowski's theory and briefly discuss the possibility of a speed-up over quantum theory in communication complexity problems.

To summarise, the five novel contributions of the current chapter are as follows:

1. The theory of Density Cubes possesses a well-defined mechanism which leads to the emergence of quantum theory – analogous to the emergence of classical physics from quantum theory via decoherence.

2. The theory of Density Cubes provides an advantage over quantum theory in a computational task based on the collision problem.

3. The axioms used to define the theory Density Cubes are insufficient to uniquely characterise it. It should hence be thought more as a framework for possible theories than a unique theory.

[4] It should be noted that all theories of interest to Barnum, Müller and Ududec do satisfy these extra constraints, and so their definition suffices for all considerations of interest in [72].

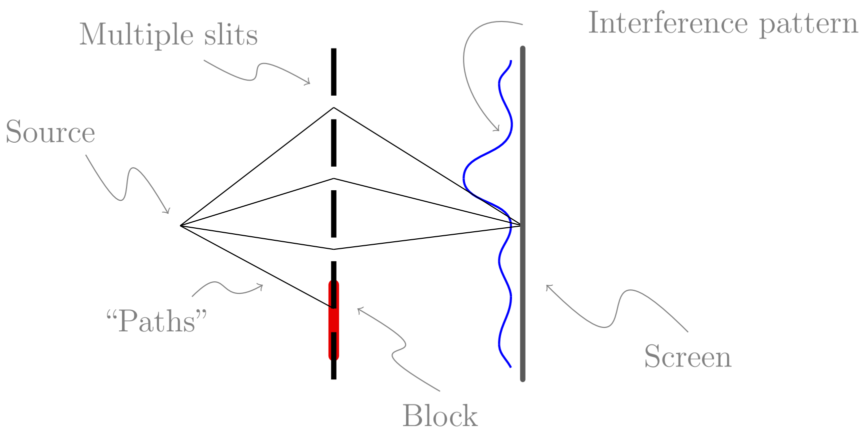


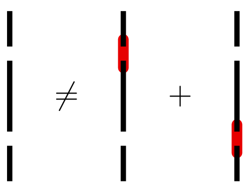
Fig. 5.1: Depiction of a multi-slit experiment

4. Quartic Quantum Theory (QQT) exhibits irreducible interference to all orders relative to the definition of higher-order interference provided by Barnum et al. in [72].

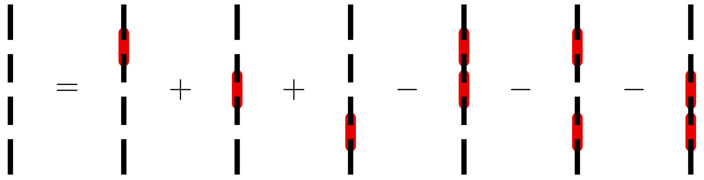
5. Point 4, above, explicitly highlights an ambiguity in the current definition of higher-order interference which must be taken into account in future experimental investigations of higher-order interference.

5.2 A definition of higher-order interference in generalised probabilistic theories

Informally, a theory is said to have nth order interference if one can generate interference patterns in an n-slit experiment, such as that depicted in Fig. 5.1, which cannot be created in any experiment with only m-slits, for all $m < n$. More precisely, this means that the interference pattern created on the screen cannot be written as a particular linear combination of the patterns generated when different subsets of slits are blocked. In the two slit experiment, quantum interference corresponds to the fact that the interference pattern cannot be written as the sum of the single slit patterns:



It was first shown by Sorkin [124, 125] that—at least for ideal experiments [152]—quantum theory is limited to the $n = 2$ case. That is, the interference pattern created in a three—or more—slit experiment *can* be written in terms of the two and one slit interference patterns obtained by blocking some of the slits. Schematically, this can be represented as:



where the role of the minus signs results is to compensate for the fact that we have overcounted each slit. If a theory does not have nth order interference then one can show it will not have mth order interference, for any $m > n$ [124].

Higher-order-interference was initially formalised by Sorkin in the framework of Quantum Measure Theory [124] but has more recently been adapted to the setting of generalised probabilistic theories in [72, 78]. Barnum, Müller and Ududec [72] have provided a definition of higher-order interference in generalised probabilistic theories which is equivalent to Sorkin's original definition in the quantum and classical cases. This definition takes its motivation from the set-up of certain experimental interference experiments, in which a quantum particle (a photon or electron, say) passes through apertures corresponding to a collection of slits in a physical barrier. By blocking some of the slits and repeating the experiment many times, one can build up an interference pattern on a screen placed behind the physical barrier. The entire physical situation is illustrated in Fig. 5.1.

The Barnum *et al.* definition of higher-order interference proceeds as follows. They firstly define *exposed faces*[5], F_i, of $\mathbf{St}(\mathbf{A})$, for any system A, as a set of states for which there exists an effect $(e_i| \in \mathbf{Eff}(\mathbf{A})$ satisfying $(e_i|s) = 1 \iff |s) \in F_i$. We should think of the effect $(e_i|$ as the effect corresponding to placing a detector just behind the slit i. The face F_i therefore corresponds to the set of states which are detected at slit i with certainty. The union of two exposed faces, $F_{ij} := F_i \cup F_j$, is defined as the smallest exposed face that includes both F_i and F_j. This is the face generated by an effect arising as a coarse graining of the effects corresponding to slit i and j respectively. Faces are disjoint $F_i \perp F_j$ if $(e_i|s) = 0$, $\forall |s) \in F_j$ and $(e_j|s) = 0$, $\forall |s) \in F_i$. We expect faces corresponding to an n-slit experiment to be disjoint; if we know with certainty that the particle has passed through a particular slit, there should be no probability of finding it at another slit. The union of multiple exposed faces $F_I := \bigcup_{i\in I} F_i$, with $I \subseteq \{1, ..., n\}$, is defined as the smallest exposed

[5] Recall that a face F is a convex set with the property that $px + (1-p)y \in F$, for any $0 \leq p \leq 1$, implies $x, y \in F$.

face that includes all F_i with $i \in I$. As before, membership in F_I corresponds to the existence of an effect $(e_I| \in \mathbf{Eff}(\mathbf{A})$ satisfying $(e_I|s) = 1 \iff |s) \in F_I$.

An n-slit experiment requires a system that has n disjoint exposed faces $F_i,\ i \in \{1, ..., n\}$. Consider an effect $(E|$ which represents the effect corresponding to the probability of finding a particle at a particular point on the screen. Then an n-slit experiment is a collection of effects $(e_I|,\ I \subseteq \{1, ..., n\}$ such that

$$(e_I|s) = (E|s), \quad \forall |s) \in F_I := \bigcup_{i \in I} F_i, \tag{5.1}$$

$$\text{and, } (e_I|s) = 0, \quad \forall |s) \text{ where } s \perp F_I. \tag{5.2}$$

We can see these effects as being the composition of the transformation induced by closing the slits $\{1, ..., n\} \setminus I$ and the effect $(E|$. If the particle was prepared in a state such that it would be unaffected by the slit closure (i.e. $|s) \in F_I$) then this composition should act the same as $(E|$ so that $(e_I|s) = (E|s)$. If instead the particle is prepared in a state which is guaranteed to be blocked (i.e. $|s') \perp F_I$) then we should not observe it, corresponding to $(e_I|s') = 0$.

The relevant quantities for the existence of various orders of interference are therefore,

$$\begin{aligned}
I_1 &:= (E|s), && (5.3)\\
I_2 &:= (E|s) - (e_1|s) - (e_2|s), && (5.4)\\
I_3 &:= (E|s) - (e_{12}|s) - (e_{23}|s) - (e_{31}|s) + (e_1|s) + (e_2|s) + (e_3|s), && (5.5)\\
I_n &:= \sum_{\emptyset \neq I \subseteq \{1,...,n\}} (-1)^{n-|I|}(e_I|s), && (5.6)
\end{aligned}$$

for some state $|s)$ and defining $(e_{\{1,...,n\}}| := (E|$. A theory has nth order interference if there exists a state $|s)$ such that $I_n \neq 0$. Lack of third-order interference therefore means that the three slit interference pattern is the sum of the two-slit patterns minus the sum of the one-slit patterns. This is what we find for quantum theory. It was shown in [124] that $I_n = 0 \implies I_{n+1} = 0$, so if we have no nth order interference then there will be no $(n + 1)$th order interference. It can be shown that classical probability theory satisfies $I_2 = 0$ and quantum theory satisfies $I_3 = 0$. The failure of $I_2 = 0$ for quantum theory means that the two-slit pattern is not the sum of the one-slit patterns, which corresponds to the usual notion of interference in the two-slit experiment.

5.3 Density cubes

Dakić *et al.* [121] have proposed a method to construct a theory that exhibits third-order interference and extends standard quantum theory. They argue, based on the results in [78], that the absence of third-order interference in quantum theory can be traced back to the fact that a quantum state coherently links at most two levels of the quantum system. This can be summarised as the fact that a quantum state is represented by a density matrix, where the matrix entries ρ_{ij}, $i \neq j$, are the coherences linking the levels i and j. So in order for a theory to exhibit third-order interference the representation of states in said theory must contain terms that coherently link *three* levels, i.e. terms of the form ρ_{ijk}, with i, j, k all distinct. Thus a potential way to construct a theory that exhibits third-order interference is to consider a theory where the states are described not by matrices ρ_{ij} as in quantum theory, but by (rank 3) *tensors* with elements of the form ρ_{ijk}. Dakić *et al.* refer to such tensors as *density cubes*, as opposed to the density matrices of quantum theory.

5.3.1 States and effects

The basic features of the theory of density cubes are defined in analogy with quantum theory, as follows[6]. Every measurement outcome is associated with a density cube[7] which, in general, has complex entries ρ_{ijk}. The element ρ_{iii} is chosen to be real and corresponds to the probability of the outcome $i = 1, ..., n$ of a particular measurement. Thus $\sum_i \rho_{iii} = 1$ and $\rho_{iii} \geq 0$. In analogy to quantum theory, we refer to this property as the *trace* of the density cube. In standard quantum theory the probability of finding the quantum state ρ in the state σ on measurement is given by $p = Tr(\rho^\dagger \sigma) = \rho^*_{ij}\sigma_{ij}$, where Einstein's summation convention has been adopted. In a similar manner, define $p = (\rho, \sigma) = \rho^*_{ijk}\sigma_{ijk}$, where p denotes the probability of finding the a density cube in state ρ when the measurement corresponding to the state σ is applied. To ensure that p is a real number, the constraint $\rho^*_{ijk}\sigma_{ijk} = \sigma^*_{ijk}\rho_{ijk}$ is enforced. In the quantum case $p \in \mathbb{R}$ is ensured as ρ_{ij} is a Hermitian matrix, hence $\rho_{ij} = \rho^*_{ji}$. Similarly, call a density cube *Hermitian* if exchanging two indices gives a complex conjugated element. As in the case of Hermitian matrices, Hermitian cubes form a real vector space with the inner product given by $(\rho, \sigma) = \rho^*_{ijk}\sigma_{ijk}$. We define pure states as those that satisfy the above conditions and also satisfy $(\rho, \rho) = 1$. Positivity of the inner product,

[6] See [121] for a more comprehensive discussion.

[7] i.e. the authors of [121] require that their theory has a one-to-one correspondence between states and effects, in terms of GPTs this means that the state and effect cones are the same.

Hermiticity and the requirement that the terms ρ_{iii} are probabilities are the only constraints imposed by [121] on the structure of density cubes, and their state space.

For a three-level system, the normalization and Hermiticity conditions imply:

1. $\rho_{iij} = \rho^*_{iij} = \rho_{iji} = \rho_{jii}, \quad i, j = 1, 2, 3,\ i \neq j,$

2. $\rho_{123} = \rho_{312} = \rho_{231} = \rho^*_{213} = \rho^*_{321} = \rho^*_{132},$

3. $\rho_{111} + \rho_{222} + \rho_{333} = 1,$

4. $\rho_{iii} \geq 0, \quad i = 1, 2, 3.$

Thus the density cube of a three-level system is specified by ten real parameters: point 1. contributes six real parameters (one for each choice of i and j), point 2. contributes one complex, or two real, parameters, point 3. contributes three real parameters and point 4. reduces by one. This is two real parameters (one complex parameter) more than what is required to specify the state of a general three-level system (qutrit) in quantum theory. Thus, the elements ρ_{ijk} with i, j, k distinct can be seen as the crucial difference between the density matrix and the density cube. Therefore, based on the results in [78] discussed above, one might naively expect that the existence of the term ρ_{ijk}, with i, j, k distinct, implies the existence of genuine third-order interference.

The complete characterisation of the density cube state space remains an important and interesting open problem. Nevertheless, some genuinely non-quantum density cubes were presented in [121]. An example of such non-quantum density cubes (i.e. those with $\rho_{123} \neq 0$) are the following three pure states, first presented in [121]:

$$\rho^{(j)} = \left\{ \begin{pmatrix} \frac{1-\delta_{1j}}{2} & 0 & 0 \\ 0 & 0 & \frac{\omega^{j-1}}{2\sqrt{3}} \\ 0 & \frac{(\omega^{j-1})^*}{2\sqrt{3}} & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & \frac{(\omega^{j-1})^*}{2\sqrt{3}} \\ 0 & \frac{1-\delta_{2j}}{2} & 0 \\ \frac{\omega^{j-1}}{2\sqrt{3}} & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \frac{\omega^{j-1}}{2\sqrt{3}} & 0 \\ \frac{(\omega^{j-1})^*}{2\sqrt{3}} & 0 & 0 \\ 0 & 0 & \frac{1-\delta_{3j}}{2} \end{pmatrix} \right\},$$

for $j = 1, 2, 3$, where $\omega = e^{i\frac{2\pi}{3}}$ and δ_{ij} is the Kronecker delta. In each of the above density cubes, the element ρ_{ijk} occurs in the jkth entry of the ith matrix in the list. It is easy to check that these density cubes are orthonormal, i.e. $(\rho_i, \rho_j) = \delta_{ij}$, and can be taken as part of a orthonormal basis in the real vector space of density cubes. We define a *physical basis* as a set of density cubes that are orthogonal and sum to $\sum_n \delta_{in}\delta_{jn}\delta_{kn}$, these physical bases correspond to allowed (pure) measurements for density cubes.

5.3.2 Transformations

An example of a genuine 'non-quantum' transformation between density cubes was also presented in [121]. In order to present the constraints on transformations between density cubes imposed in [121], consider the following. Take the complex vector space of general rank-3 tensors, the Hermitian cubes, defined above, form a real subspace within this. A complex subspace can be defined by Span$[C^{(i)}]$ where $C^{(i)}$ are defined as,

$$C^{(n)}_{ijk} = \delta_{in}\delta_{jn}\delta_{kn}, \quad n = 1, 2, 3,$$

$$C^{(k)} = \frac{1}{\sqrt{3}}\left\{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \delta_{4k} \\ 0 & \delta_{5k} & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & \delta_{5k} \\ 0 & 0 & 0 \\ \delta_{4k} & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \delta_{4k} & 0 \\ \delta_{5k} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right\}, \quad k = 4, 5,$$

note that $C^{(4)}$ and $C^{(5)}$ are not Hermitian cubes[8] but the others are. A vector in Span$[C^{(i)}]$ is specified by five complex numbers. If we take the intersection of Span$[C^{(i)}]$ with the Hermitian cubes we obtain another real vector subspace where in the $C^{(i)}$ basis[9] vectors are of the form $(p_1, p_2, p_3, z, z^*)^T$, $p_i \in [0, 1] \subset \mathbb{R}_+$, $z \in \mathbb{C}$ and with $\sum_{i=1}^{3} p_i = 1$. This is a subspace of the Hermitian cubes. We must also impose our constraints as before, which gives the state space as a convex set living in this subspace.

The authors of [121] consider only transformations that leave this subspace invariant. Aside from this the *only* requirements imposed by the authors of [121] are that the transformations are unitary matrices that map at least one physical basis of density cubes to another physical basis.

For example, consider a unitary transformation $T : D_0 \to D$, where $D_0 = \{q_1, q_2, q_3\}$ and $D = \{\rho_1, \rho_2, \rho_3\}$ are defined (in the $C^{(i)}$ basis) as follows,

$$\begin{aligned} q_1 &= (1, 0, 0, 0, 0)^T, & \rho_1 &= \frac{1}{2}(0, 1, 1, 1, 1)^T, \\ q_2 &= (0, 1, 0, 0, 0)^T, & \rho_2 &= \frac{1}{2}(1, 0, 1, \omega, \omega^*)^T, \\ q_3 &= (0, 0, 1, 0, 0)^T, & \rho_3 &= \frac{1}{2}(1, 1, 0, \omega^*, \omega)^T, \end{aligned}$$

[8] A similar situation occurs in quantum theory: the Pauli matrices form a basis of the real vector space of Hermitian matrices, yet individual Pauli matrices are not physical states, only certain linear combinations of them are.

[9] We could instead use the basis which uses $C^{(4)} + C^{(5)}$ and $C^{(4)} - iC^{(5)}$ in which case our vectors would be written as five real numbers.

where as before $\omega = e^{\frac{2\pi i}{3}}$.

The q_i's span a subspace of the 'quantum states' of these density cubes. One matrix, provided by Dakić *et al.*, that satisfies the conditions $Tq_i = \rho_i$, leaves this subspace invariant and is unitary is,

$$T = \frac{1}{2}\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & \omega^* & \omega \\ 1 & 1 & 0 & \omega & \omega^* \\ 1 & \omega & \omega^* & 1 & 0 \\ 1 & \omega^* & \omega & 0 & 1 \end{pmatrix}. \tag{5.7}$$

Note that there are many matrices that satisfy the above condition, see [121] for a more in-depth discussion.

Introducing an appropriate post-measurement state update rule for density cubes and using the transformation from equation 5.7, it is shown in [121] that the theory of Density Cubes exhibits third-order interference, i.e. $I_3 \neq 0$ for this theory. We will not go through the details of the calculation here, but defer the reader to the original paper [121]. In section 5.4.2 however, we give the full details of the calculation showing the Quartic Quantum Theory of Życzkowski [122] exhibits higher-order interference relative to the definition provided in section 5.2.

5.3.3 Hyper-decoherence

A hyper-decoherence mechanism will now be shown to exist in the theory of Density Cubes—provided that there is an inner product preserving embedding (i.e. an injective, linear map) of the quantum states into the density cube state space.

Such an embedding was given in [121] and can be defined as follows. Denote an arbitrary quantum state by $\rho_{QT} \in \Omega_{QT}$ and an arbitrary density cube by $\rho_{DC} \in \Omega_{DC}$, where Ω_{QT} is the quantum state space, and so on. Define the embedding map $\mathcal{E}\colon \Omega_{QT} \to \Omega_{DC}$ by:

$$\big(\mathcal{E}[\rho_{QT}]\big)_{iii} = (\rho_{QT})_{ii},\ \big(\mathcal{E}[\rho_{QT}]\big)_{iij} = \sqrt{\frac{2}{3}}\mathrm{Re}(\rho_{QT})_{ij},\ \big(\mathcal{E}[\rho_{QT}]\big)_{ijj} = \sqrt{\frac{2}{3}}\mathrm{Im}(\rho_{QT})_{ij} \text{ for } i < j,$$

$$\big(\mathcal{E}[\rho_{QT}]\big)_{ijj} = -\sqrt{\frac{2}{3}}\mathrm{Im}(\rho_{QT})_{ij} \text{ for } i > j, \text{ and } \big(\mathcal{E}[\rho_{QT}]\big)_{ijk} = 0, \text{ for } i \neq j \neq k \neq i.$$

The other elements of the density cube are defined by the Hermiticity condition described in section 5.3.1 One can check that this embedding preserves the inner product. That is, we have that

$$(\rho_{QT}, \sigma_{QT})_{QT} = (\mathcal{E}[\rho_{QT}], \mathcal{E}[\sigma_{QT}])_{DC},$$

where $(.,.)_{QT}$ is the inner product between quantum states, and so on.

To discuss hyper-decoherence it is useful to separate the density cubes into third order and lower order terms, we therefore write a generic density cube as,

$$\rho_{DC} = \rho_{DC}^{(3)} + \rho_{DC}^{(2,1)}$$

where we define,

$$(\rho_{DC}^{(3)})_{ijk} := \begin{cases} (\rho_{DC})_{ijk} & \text{if } i \neq j \neq k \neq i \\ 0 & \text{otherwise} \end{cases},$$
$$(\rho_{DC}^{(2,1)})_{ijk} := \begin{cases} 0 & \text{if } i \neq j \neq k \neq i \\ (\rho_{DC})_{ijk} & \text{otherwise} \end{cases}.$$

Note that $\rho_{DC}^{(2,1)}$ and $\rho_{DC}^{(3)}$ are not necessarily themselves valid density cubes.

Given the above embedding, $\mathcal{E}$, one can define a hyper-decoherence map $\mathcal{D}$ as follows:

$$\mathcal{D} \circ \mathcal{E} = \mathbb{I}_{QT}, \quad \mathcal{D}[\rho_{DC}^{(3)}] = 0,$$

where $\mathcal{D}$ is a linear map[10] from the real vector space of Hermitian cubes to the real vector space of Hermitian matrices, and $\mathbb{I}_{QT}$ is the identity transformation on Hermitian matrices. This choice of $\mathcal{D}$ seems natural as we would expect such a map to leave any quantum state embedded in the Density Cube state space invariant and to eliminate the higher order coherences.

In order to show that $\mathcal{D}$ is a valid hyper-decoherence map we need to show that it maps all density Cube states to valid quantum states. That is $\mathcal{D}[\rho_{DC}]$ must be a positive, Hermitian operator with unit trace. That $\mathcal{D}[\rho_{DC}]$ has unit trace is guaranteed by the definition of $\mathcal{D}$ and the construction of the Density Cubes. To check the Hermiticity condition, consider the following. From the definition of $\mathcal{E}$ given above we can write $(\mathcal{D}[\rho_{DC}])_{ij}$ as:

$$\begin{aligned} (\mathcal{D}[\rho_{DC}])_{ij} &= \sqrt{\frac{3}{2}}\Big(\big(\mathcal{E} \circ \mathcal{D}[\rho_{DC}]\big)_{iij} + i\big(\mathcal{E} \circ \mathcal{D}[\rho_{DC}]\big)_{ijj} \Big), \text{ for } i < j \\ \text{and, } (\mathcal{D}[\rho_{DC}])_{ij} &= \sqrt{\frac{3}{2}}\Big(\big(\mathcal{E} \circ \mathcal{D}[\rho_{DC}]\big)_{iij} - i\big(\mathcal{E} \circ \mathcal{D}[\rho_{DC}]\big)_{ijj} \Big), \text{ for } i > j. \end{aligned} \tag{5.8}$$

To show $\mathcal{D}[\rho_{DC}]^\dagger = \mathcal{D}[\rho_{DC}]$, we must check that $\big(\mathcal{D}[\rho_{DC}]\big)_{ij} = \big(\mathcal{D}[\rho_{DC}]\big)^*_{ji}$ for all i, j, but this follows from applying the Density Cube Hermiticity condition to equations (5.8).

[10] Note that linearity ensures that we can extend the map from the states on which it is defined to all Hermitian cubes.

To check the positivity property, we need to show that $\big(\mathcal{D}[\rho_{DC}],\sigma_{QT}\big)_{QT}\geq 0$, for all ρ_{DC} and σ_{QT}. Note that we can always choose suitable real coefficients c_i such that $\rho^{(2,1)}_{DC}=\sum_i c_i\mathcal{E}[\rho^i_{QT}]$, where $\rho^i_{QT}\in\Omega_{QT}$ are some arbitrary set of density matrices. We can therefore write an arbitrary density cube as $\rho_{DC}=\sum_i c_i\mathcal{E}[\rho^i_{QT}]+\rho^{(3)}_{DC}$. Combining this with the definition of $\mathcal{D}$, we have that

$$\begin{aligned}
\big(\mathcal{D}[\rho_{DC}],\sigma_{QT}\big)_{QT} =&\big(\mathcal{E}\circ\mathcal{D}[\rho_{DC}],\mathcal{E}[\sigma_{QT}]\big)_{DC}\\
=&\big(\mathcal{E}\circ\mathcal{D}\big[\sum_i c_i\mathcal{E}[\rho^i_{QT}]+\rho^{(3)}_{DC}\big],\mathcal{E}[\sigma_{QT}]\big)_{DC}\\
=&\big(\sum_i c_i\mathcal{E}[\rho^i_{QT}],\mathcal{E}[\sigma_{QT}]\big)_{DC}\\
=&\big(\rho_{DC},\mathcal{E}[\sigma_{QT}]\big)_{DC}-\big(\rho^{(3)}_{DC},\mathcal{E}[\sigma_{QT}]\big)_{DC}\\
=&\big(\rho_{DC},\mathcal{E}[\sigma_{QT}]\big)_{DC}\\
\geq&0,
\end{aligned}$$

where $\big(\rho^{(3)}_{DC},\mathcal{E}[\sigma_{QT}]\big)_{DC}=0$ follows from $(\mathcal{E}[\sigma_{QT}])_{ijk}=0$ for $i\neq j\neq k\neq i$. The equation $\big(\mathcal{D}[\rho_{DC}],\sigma_{QT}\big)_{QT}=\big(\rho_{DC},\mathcal{E}[\sigma_{QT}]\big)_{DC}$, derived above, implies that the embedding map $\mathcal{E}$ is the adjoint of the hyper-decoherence map $\mathcal{D}$. This may prove useful in further constructions of higher-order interference theories.

Given the embedding $\mathcal{E}$, the hyper-decoherence map defined above maps density cubes to valid quantum states. One should note however, that the existence of this embedding is not guaranteed by the axioms of the Density Cube framework, but is a very reasonable constraint if one wants an extension of quantum theory.

In quantum theory, to have coherence between two levels of the quantum state described by the density matrix ρ_{ij} there must be some probability of finding the state in either of the levels that the coherence is between. These probabilities set a bound on the degree of coherence possible, e.g. for a qubit we have $|\rho_{01}|^2\leq\rho_{00}\rho_{11}$. Based on this, one might expect that any third order coherence in a Density Cube would be supported by second and first order coherences. Interestingly this is not the case in the Density Cube framework; states in the physical basis D considered by Dakíc *et al.* have third order terms but all second order terms are zero. While it is the case that the positivity condition imposes some bounds on the higher-order coherences, there may be further constraints that need to be imposed to have a well-defined theory. It would be interesting if future constructions of higher-order interference theories had this property.

5.3.4 A computational advantage?

When comparing quantum theory with other foil theories, an approach that has proved fruitful in recent years is to compare their performance in information-theoretic tasks. We will now show that the theory of Density Cubes has a slight advantage over quantum theory in a computational task we call the 'three collision problem', which is a variation of the standard collision problem discussed in [133]. The three collision problem is defined as follows: given a function from a trit to a bit, $f : \{0, 1, 2\} \to \{0, 1\}$, determine if $f(0) = f(1) = f(2)$. As is standard in quantum computation, we represent this problem with a black-box oracle. Performance will be measured via the probability of error after a single query to this oracle, given the caveat that if $f(0) = f(1) = f(2)$ there must be zero error.

Let $\{|i\rangle\}$, for $i = 0, 1, 2$, be the quantum computational basis and consider the following quantum oracle for this problem:

$$\mathcal{O}_f^{QT}|i\rangle = (-1)^{f(i)}|i\rangle.$$

This oracle is the same as the one considered by Grover in his search algorithm [1], and it is easy to check it is unitary. Preparing a superposition over the three basis states and querying the oracle leaves us in the state

$$\frac{1}{\sqrt{3}}\big((-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle + (-1)^{f(2)}|2\rangle\big).$$

If $f(0) = f(1) = f(2)$, then the state, up to a global phase, is: $\frac{1}{\sqrt{3}}\big(|0\rangle + |1\rangle + |2\rangle\big)$, while if they are not equal the state, up to a global phase, is one of: $\frac{1}{\sqrt{3}}\big(-|0\rangle + |1\rangle + |2\rangle\big)$, $\frac{1}{\sqrt{3}}\big(|0\rangle - |1\rangle + |2\rangle\big)$, or $\frac{1}{\sqrt{3}}\big(|0\rangle + |1\rangle - |2\rangle\big)$. As the state with $f(0) = f(1) = f(2)$ is not orthogonal to the other three, there does not exist a measurement that can perfectly distinguish them and the error after one query is 1/9.

Dakíc *et al.* have provided a way to associate one of three density cubes to a pure three-level quantum state $|\psi\rangle = c_0|0\rangle + c_1|1\rangle + c_2|2\rangle$. The association is as follows:

$$\rho^{(n)}_{iij} = -\frac{1}{\sqrt{6}}\mathrm{Re}(c_i^* c_j), \quad \rho^{(n)}_{ijj} = -\frac{1}{\sqrt{6}}\mathrm{Im}(c_i^* c_j) \text{ for } i < j, \quad \rho^{(n)}_{iii} = \frac{1}{2}(1 - |c_i|^2),$$
$$\text{and,} \quad \rho^{(n)}_{012} = \frac{\omega^n}{2\sqrt{3}} \text{ with } n = 0, 1, 2.$$

The other elements of the density cube are determined by the Hermiticity condition (see section 5.3.1). Note that only the third-order terms depend on the value of n. One can show [121] that $(\rho^{(n)}(|\phi\rangle), \rho^{(m)}(|\psi\rangle))_{DC} = \frac{1}{4}(1 + |\langle\phi|\psi\rangle|^2) + \frac{1}{2}\cos\frac{2\pi(n-m)}{3} \geq 0$. Given this association, we can describe a Density Cube oracle for the three collision problem as follows. The oracle acts as the quantum oracle on the 'quantum part' of the density cube, but also acts on the 'higher order term' i.e. the value of n. We define the oracle as,

$$\mathcal{O}_f^{DC} :: \rho^{(n)}(|\psi\rangle) \mapsto \rho^{(n_f)}(\mathcal{O}_f^{QT}|\psi\rangle)$$

where $n_f = n + f(0) + f(1) + f(2) = n + \sum_i f(i)$. One can check that the action of this oracle leaves the fragment given by the above association invariant. While it may appear odd at first to allow density cubes with non-zero higher-order terms access to the value $f(0) + f(1) + f(2)$, it should be noted that quantum theory has a similar advantage over classical theory when accessing a computational oracle. In classical computing one can only access the value of f on a single value i per query of an oracle, but, in quantum theory, one can access information about $f(i) + f(j)$ by querying the same oracle in superposition. It thus seems reasonable to allow third-order terms $\rho_{012}^{(n)}$ access to information about the value of $f(0) + f(1) + f(2)$.

Let $|\phi\rangle = \frac{1}{\sqrt{3}}\big(|0\rangle + |1\rangle + |2\rangle\big)$, and prepare the density cube $\rho^{(0)}(|\phi\rangle)$. Applying the Density Cube oracle leaves this state invariant if $f(0) = f(1) = f(2)$ and maps this state to either $\rho^{(1)}(\mathcal{O}_f^{QT}|\phi\rangle)$ or $\rho^{(2)}(\mathcal{O}_f^{QT}|\phi\rangle)$ otherwise, thus giving an error probability of

$$\Big(\rho^{(0)}\left(|\phi\rangle\right), \rho^{(1)}(\mathcal{O}_f^{QT}|\phi\rangle)\Big) = \Big(\rho^{(0)}(|\phi\rangle), \rho^{(2)}(\mathcal{O}_f^{QT}|\phi\rangle)\Big) = 1/32$$

after a single query. The theory of Density Cubes thus provides a slight advantage over quantum theory in the three collision problem.

5.3.5 Issues with the Density Cube framework

In this section two possible issues with the framework of Density Cubes will be presented and discussed. In particular it will be demonstrated that the axioms imposed in defining the theory are insufficient to uniquely characterise the state space. We also show that the definition of transformations employed by [121] allows for transformations in the theory that map well-defined states to density cubes that give complex-valued probabilities for certain measurement outcomes.

5.3.5.1 Axioms insufficient to specify a unique operational theory

Dakić *et al.* mention that they have not fully constructed the state space for density cubes [121], instead they present a particular set of states which satisfy their axioms (i.e. they are Hermitian, have unit trace and are each positive with respect to the others). The difficulty in fully constructing the state space stems from the positivity axiom. In quantum theory we can define positivity as,

$$Tr(\rho^\dagger\sigma) \geq 0 \quad \forall \rho, \sigma,$$

where ρ and σ are density matrices. This is analogous to the positivity condition imposed by Dakić *et al.* and we refer to this property as 'relative positivity'. In practice this is a difficult property to use to construct a state space, there is—potentially—an infinite number of conditions to check for each state in the state space. In quantum theory we can avoid this problem by using an alternative—and equivalent—definition of positivity, that

$$\forall \lambda \in \mathsf{Eigenvalues}(\rho) \quad \lambda \geq 0.$$

This is a single state property rather than a relative property and so it is simple to construct a state space by imposing this condition. As we do not have a notion of eigenvalues for rank 3 tensors, there is no equivalent condition for density cubes. We are therefore limited to using relative positivity.

Given that we only have a relative notion of positivity, it is possible to construct different state spaces depending on which set of states we choose to start with. However we know that—if we want a genuine extension of quantum theory—we need some (Hermitian, trace and inner product preserving) embedding of the quantum states into the Density Cube state space. Dakić *et al.* present one such embedding, which we discussed in section 5.3.3. One hope is that, once we restrict to such an embedding, the state space becomes uniquely specified. Moreover, one would hope that that this choice of embedding is analogous to a of re-parametrization of the quantum state space, and, as such, leads to operationally equivalent theories. Unfortunately this is not the case; it is possible to construct operationally distinct theories within the Density Cube framework. As such, the axioms imposed are not sufficient to uniquely characterise the theory.

For example, consider the embedding of quantum states described in [121], discussed in section 5.3.3, and use the basis $\{C^{(i)}\}$ described above. Then we can consider the states

$$c = \frac{1}{2}(1,1,0,1,1)^T \text{ and } v = \frac{1}{256}\left(10,10,236,-\left(65+i\sqrt{595}\right),-\left(65-i\sqrt{595}\right)\right)^T,$$

these are both quantum states with added higher-order coherence terms and so will be positive with respect to all of the quantum states. However they are not positive with respect to each other, $(c,v) < 0$. There is no reason to prefer one of these to the other and we cannot add both to the state space, we therefore have an arbitrary choice at this stage in how to construct the theory. Note that both of these states are positive with respect to the physical basis D, but have different inner products with elements of D. Hence choosing which state to include will lead to theories that make operationally distinct predictions about certain measurements.

Given the above discussion it may therefore be better to consider the theory of density cubes more as a framework for developing theories in. The (partial) state space of Dakić *et al.* would then be one example of a state space within this framework. The difficulty in constructing the complete state space causes further problems when defining transformations within the theory. In theories where there is a complete geometric view of the state space, as was the case for Boxworld in chapter 1, it is simple to define transformations as linear transformations that map the state space into itself. However, if we are not given a complete state space it is not possible to define transformations in this way.

5.3.5.2 Characterising the set of physical transformations

We will now show that the lack of fully constructed state space is also problematic for defining allowed transformations within the theory. Dakić *et al.* present a particular transformation T that they use throughout their paper. It can be shown that for the particular fragment they are considering, this is a valid transformation. By valid transformation we mean that it is linear and maps states to states. They also provide a set of axioms which need to be satisfied such that a transformation is valid. These conditions turn out to be necessary but not sufficient, as we shall now demonstrate.

The conditions imposed on transformations in [121]—as discussed in section 5.3.2—are as follows:

1. linearity
2. unitarity
3. subspace preserving
4. map between physical bases (e.g. $D_0 \mapsto D$).

This allows for transformations such as T', equation 5.9, which can easily be shown to violate the Hermiticity of states.

$$T' = \frac{1}{2}\begin{pmatrix} 0 & 1 & 1 & \frac{1}{2}(1+\sqrt{3}) & \frac{1}{2}(-1+\sqrt{3}) \\ 1 & 0 & 1 & \frac{1}{2}(\omega^* + \sqrt{3}\omega) & \frac{1}{2}(-\omega + \sqrt{3}\omega^*) \\ 1 & 1 & 0 & \frac{1}{2}(\omega + \sqrt{3}\omega^*) & \frac{1}{2}(-\omega^* + \sqrt{3}\omega) \\ 1 & \omega & \omega^* & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 1 & \omega^* & \omega & \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}. \tag{5.9}$$

For example, $T'(\rho_1)$ has complex entries where real ones should be, and so gives complex-valued "probabilities" on some measurements.

$$\begin{aligned} T'(\rho_1) =& \left(\frac{1}{2} + \frac{\sqrt{3}}{4}, \frac{1}{4}\left(1 - \sqrt{3}\left(\frac{1+i}{2}\right)\right), \frac{1}{4}\left(1 - \sqrt{3}\left(\frac{1-i}{2}\right)\right), \frac{\sqrt{3}-1}{8}, \frac{\sqrt{3}-3}{8}\right)^T \\ \approx& (0.9, 0.03 - 0.2i, 0.03 + 0.2i, 0.09, -0.2)^T. \end{aligned}$$

The usual solution to this would be to require that transformations map states to states or equivalently that they preserve Hermiticity and positivity, which would rule out unphysical transformations such as T'.

Using Hermiticity and positivity preservation as a characterisation of transformations however is dependent on the state space, and, as we do not have a complete state space, these are impossible to enforce in practice. Characterising transformations beyond the specific example of Dakić *et al.* is not possible at this stage.

This again highlights the issue that different fragments give operationally distinct predictions, we see here that not only the possible states depend on the choice of state space, but that the set of physical transformations does as well.

5.4 Quartic quantum theory

Quartic quantum theory (QQT) was developed by Życzkowski [122] as an attempt to realise the $K = N^4$ level of the tomographic hierarchy introduced by Hardy in [18]. This is to be contrasted to quantum theory which satisfies[11] $K = N^2$ and classical probability theory which satisfies $K = N$. Density cubes however satisfy $K = N^2 + 2\binom{N}{3} \neq N^r$ and so are not in Hardy's hierarchy[12].

We have discussed the connection between tomography and higher-order interference presented in [78], specifically, the dimension of the subspace on which one

[11] This is allowing for sub-normalised states hence quantum theory having $K = N^2$ rather than $K = N^2 - 1$.

[12] Thus implying that the theory of Density Cubes violates tomographic locality.

must perform measurements to do complete tomography corresponds to the order of interference. In the $K = N^r$ hierarchy, post-quantum theories require tomography on greater than two dimensional subspaces. In light of this, QQT provides a potential candidate for a GPT that exhibits higher order interference.

5.4.1 Description of the theory

We will now provide a brief overview of the theory (see the original paper [122] for a more in-depth discussion). The state space for an N level QQT system is constructed from a restriction of an N^2 level quantum system (i.e. the tensor product of two N level quantum systems). The restriction limits us to convex combinations of states that are unitarily connected to[13] $|s_{initial}) := \frac{1}{N}\mathbb{I} \otimes |0\rangle \langle 0|$, so the state space is given by the convex hull of $|s) \in \{U(\frac{1}{N}\mathbb{I} \otimes |0\rangle \langle 0|)U^\dagger \mid U \in SU(N^2)\}$, i.e. we allow for arbitrary mixtures of any states which can be reached by applying arbitrary unitaries to the composite quantum system beginning in state $|s_{initial})$.

This restriction on the quantum state space implies there is a maximum purity any state can reach. This can be seen to be roughly analogous to the epistemic restriction used by Spekkens in his Toy Model [123], where the state space of a two level system is given by a pair of classical bits (i.e. a pair of two level classical systems) with a restriction imposed on how much one can know about the system.

Transformations are defined as linear maps that leave the state space invariant and which are completely preserving, i.e. $T : \Omega_N \to \Omega_N$ and $T \otimes \mathbb{I}_M : \Omega_{NM} \to \Omega_{NM} \;\; \forall M$, where Ω_N is the QQT state space for an N level system. The last condition is a generalisation of complete positivity in quantum theory.

Effects satisfy the 'no-restriction hypothesis' [126] which says that any mathematically well-defined effect is allowed. That is an effect $(e|$ is allowed in the theory if it is linear and $0 \leq (e|s) \leq 1$, $\forall |s) \in \Omega_N$. We have imposed a restriction on the Quantum theory state space[14], and so the effect space is enlarged. For example we can have effects such as $(e| = N |0\rangle \langle 0| \otimes |0\rangle \langle 0|$. Which in quantum theory could give probabilities greater than one, but due to the restriction on purity this cannot happen in QQT. This is because $(e|s) \leq N\lambda_s^{max}$. Where λ_s^{max} is the maximum eigenvalue of

[13] Where we are using curved brackets to denote QQT states and effects and Dirac brakets to represent the underlying quantum density matrix description of the state

[14] Geometrically the un-normalised state space of any causal theory is a convex cone [15] (the normalised state space is the intersection of a hyperplane with the convex cone), and if the no-restriction hypothesis is satisfied then the effect space is the dual cone. If the state space is restricted this increases the size of the dual cone, which is what we find in QQT.

the density matrix representation of the state $|s)$. For a QQT state $\lambda_s^{max} \leq 1/N$ so $(e|S) \leq N/N = 1$.

This fully constructs the theory for an N level quartic quantum system as we have a complete consistent description of all of the states, transformations and effects that exist in the theory. There is also a consistent notion of hyper-decoherence by which any quartic quantum system can decohere to a quantum system. In QQT decoherence is represented by a partial trace over one of the quantum sub-systems, this clearly can only map us to the quantum state space, additionally any quantum state, ρ, can be reached in this way through $\rho = Tr_{A2}(\rho \otimes \frac{1}{N}\mathbb{I}_{A2})$. It is worth reiterating that these are not physical subsystems. The choice of tensor product decomposition and which part to trace out is therefore entirely arbitrary.

5.4.2 Interference in Quartic Quantum Theory

We will consider interference in the context of the definitions described in section 5.2, and show that QQT exhibits Nth order interference in any N level system, for all N. Firstly we define the faces,

$$F_i := \left\{ \frac{1}{N} |i\rangle \langle i| \otimes \mathbb{I} \right\} = \left\{ \frac{1}{N} \sum_{j=1}^{N} |ij\rangle \langle ij| \right\},$$

we can then choose a set of effects which satisfy the constraints given by the definition. These are,

$$\begin{aligned} (E| &:= \sum_{i,j=1}^{N} |ij\rangle \langle ij|, && (5.10) \\ (e_i| &:= N\,|ii\rangle \langle ii|, \\ (e_I| &:= \sum_{i\in I} \sum_{j=1}^{N} |ij\rangle \langle ij|, \text{ for } I \subseteq \{1, ..., N\}. \end{aligned}$$

It is simple to show that these do satisfy equation 5.1 and equation 5.2 For instance if $|s) \in F_i$, we have

$$(e_i|s) = \mathrm{Tr}((N\,|ii\rangle \langle ii|)(\frac{1}{N} |i\rangle \langle i| \otimes \mathbb{I})) = 1 = (E|s)$$

and if $|s') \perp F_i$, we have

$$(e_i|s') = \mathrm{Tr}((N\,|ii\rangle \langle ii|)(\frac{1}{N} |j\rangle \langle j| \otimes \mathbb{I})) = 0,$$

as required.

Consider a three level system, $N = 3$, we have third-order interference if

$$(E| \neq \sum_{i>j} (e_{\{i,j\}}| - \sum_{i} (e_i|,$$

which is the case here. We have:

$$(E| = \sum_{i,j=1}^{N} |ij\rangle \langle ij| \neq 2 \sum_{i \neq j} |ij\rangle \langle ij| - \sum_{i} |ii\rangle \langle ii| = \sum_{i>j} (e_{\{i,j\}}| - \sum_{i} (e_i|$$

If instead of $(e_i|$ we used the effects $(e_I|$ where $I = \{i\}$, these also satisfy the conditions in the definition but don't give us third-order interference, that is

$$(E| = \sum_{i,j=1}^{N} |ij\rangle \langle ij| = \sum_{i>j} (e_{\{i,j\}}| - \sum_{i} (e_{\{i\}}|.$$

So we see that we obtain higher order interference by using the super-quantum effects allowed in QQT.

It can be shown that this approach generalises to Nth order interference. That is that we can find a set of effects such that

$$(E| \neq \sum_{\emptyset \neq I \subseteq \{1,\ldots,N\}} (-1)^{N-|I|} (e_I|.$$

A valid set of effects for this are those defined above in equation 5.10, where we see Nth order interference if we replace $(e_{\{i\}}| \to (e_i|$. The simplest way to see this is to observe that the effects $(e_I|$ are all the quantum effects for a N-slit experiment tensored with the identity, therefore using these we will not see higher-order interference, but if we replace the 'quantum' $(e_{\{i\}}|$ with a super-quantum $(e_i|$ and note that $\sum_i (e_i| \neq \sum_i (e_{\{i\}}|$ then we will see higher order interference to all orders.

Note that we obtain this result as the constraints imposed in the definition of higher-order interference are insufficient to uniquely determine the effects $(e_I|$, this is perhaps a problem with the definition. If one were to actually perform the experiment then there should be a unique description of the effect corresponding to what happens, as, at the operational level, it should arise from blocking slits in a physical barrier. Thus for the definition of Barnum *et al.* to correspond to this physical picture in an operationally meaningful way, extra constraints must be imposed on the theory under consideration. Note that, such extra constraints are present in all situations of interest to Barnum *et al.* in [72].

Based on the above discussion, one should not consider QQT as an example of a theory that exhibits higher-order interference in the sense originally meant by Sorkin, but rather a demonstration of the challenges of applying his original definition to arbitrary theories. Thus to begin to understand the reason why, in some sense, quantum theory is limited to second-order interference, we first need a definition of higher-order interference that is applicable to, and makes good operational sense in, arbitrary generalised probabilistic theories. Such a definition will be presented in chapter 6.

5.4.3 Composite systems in Quartic Quantum Theory

The main limitation of quartic quantum theory —as discussed by Życzkowski [122]— is that it does not deal with composite systems. The difficulty with defining composite systems is ensuring that discarding part of a composite system does not result in a state outside the (single system) QQT state space. For example if we define composition in the same way as quantum theory then a bipartite QQT system would be made of four quantum systems, two of which are required to be in the maximally mixed state. If we then allow for arbitrary quantum transformations on this system then we can use a swap unitary to put all of the mixed systems into one half of the bipartition and all of the pure systems into the other. If we then discard the mixed partition we are left with a pure quantum state, which is not a valid state in QQT. In other words, marginalisation takes us outside the QQT state space.

For example, if we prepare the state $|s_{AB}) = \frac{1}{N^2} |0\rangle\langle 0|_{A1} \otimes \mathbb{I}_{A2} \otimes |0\rangle\langle 0|_{B1} \otimes \mathbb{I}_{B2}$, apply a swap to the middle two systems ($A2$ and $B1$), $U_{swap}^{A2,B1}|s_{AB}) = \frac{1}{N^2} |0\rangle\langle 0|_{A1} \otimes |0\rangle\langle 0|_{A2} \otimes \mathbb{I}_{B1} \otimes \mathbb{I}_{B2}$, then discarding system B gives, $|0\rangle\langle 0|_{A1} \otimes |0\rangle\langle 0|_{A2}$, which is outside the state space as it is 'too pure'.

A possible solution to this problem is to impose a restriction on the allowed transformations to try to avoid a situation like this. For example, allowing only separable transformations would mean that it was impossible to apply the swap between the two QQT systems and so discarding one of them could not cause problems. This would mean there are no entangling dynamics in the theory[15] and moreover that we are unable to reversibly prepare an entangled state from a product state. An interesting direction to pursue would be whether this can be seen as a consequence of third-order interference, or whether it is possible to have a theory with third-order interference and similar entangling dynamics to those that we have in quantum theory.

[15] Note that Boxworld also shares this feature [7].

5.4.3.1 Note on Boxworld-like correlations in Quartic Quantum Theory

In using the quantum tensor product in the previous section we are relying on a commonly used axiom in quantum reconstructions, that any N level system should be equivalent, i.e. a single system with N-levels should be equivalent to a composite system that has N-levels. If we relax this assumption then we can instead use some other tensor product[16].

We note that if we consider the 'classical' subspace of a two-level quartic quantum system, i.e. the diagonal density matrices, the state space corresponding to these states forms an octahedron [122], and the effect space dual to this forms a cube. This is the 'unrestricted Spekkens Toy Model' state and effect space discussed in [126]. Janotta and Lal discuss how the (generalised) maximal tensor product of such a state space gives rise to PR box correlations, i.e. those that maximally violate a Bell inequality whilst maintaining no-signalling. We therefore should be able to obtain the same correlations if we take the maximal tensor product of two two-level quartic quantum systems. Such correlations imply a speed-up over quantum theory in communication complexity problems and this opens the door to investigations of the information processing power of well-defined physical theories with higher-order interference.

We have seen that QQT is a well-defined extension of quantum theory and so may prove a useful foil in understanding the certain features of the quantum formalism.

5.5 Conclusion

This chapter provided an overview of the literature on higher-order interference and investigated two proposed extensions of quantum theory: Dakić *et al.*'s Density Cubes and Życzkowski's Quartic Quantum Theory. We examined the order of interference in these theories relative to the hierarchy defined by Sorkin, and investigated whether they satisfied natural physical conditions one would expect from an extension of quantum theory.

The specific partial state space and single transformation presented in the original Density Cube paper [121] exhibits third-order interference. However this state space is not uniquely specified by the imposed axioms, and there exist other transformations allowed by these axioms which lead to unphysical results. It would therefore be interesting to investigate what further axioms would be necessary to uniquely specify

[16] This tensor product will have to give a state space bound by the minimal and maximal tensor products, see [126] for details.

a state space, as such a construction would provide a natural way of characterising the physical transformations. We showed that if one has an embedding of quantum theory into a specific Density Cube state space, the adjoint of this embedding gives a suitable hyper-decoherence mechanism. Considering further consistency requirements with quantum theory may help with fully developing the theory and may provide a complete axiomatisation.

The operational definition of higher-order interference of Barnum *et al.* suffers from an ambiguity: the specification of the effect $(E|$ does not uniquely fix the effects $(e_I|$ in an arbitrary theory[17]. We would intuitively expect that once $(E|$ is specified the effects $(e_I|$ are fixed, as they should arise from blocking a certain number of slits in a physical barrier. Thus to begin to illuminate *why* quantum theory is limited to only second-order interference, we first need a definition of higher-order interference that is applicable to, and makes good operational sense in, arbitrary theories. This forms the starting point of the next chapter.

In quantum theory there is an intimate relation between interference and phase, which is illustrated most clearly by the Mach-Zender interferometer. The connection between phase and interference is not touched on by the Barnum *et al.* notion of higher-order interference. Garner *et al.* [130] have proposed a definition of phase and interference applicable to an arbitrary generalised probabilistic theory, but their definition of interference bears no resemblance to Sorkin's hierarchy, and as such they do not discuss higher-order interference. The subject of phase transformations and higher-order interference will be investigated in the next chapter.

It was shown in [84] that quantum interference is a sufficient condition for a quantum computer to be hard to classically simulate. In the present chapter we saw that the theory of Density Cubes provides a slight advantage over quantum theory in a certain computational task. This raises the question of whether higher-order interference is a general resource for post-quantum computation and information processing. The connection between higher-order interference and computation will be touched upon in the next chapter, and forms the main focus of chapter 7.

[17] If the theory in question supports *filters* [72], then the effects can be uniquely specified by a choice of filters. As this is the only situation of interest to Barnum *et al.*, their definition suffices for all considerations of interest in [72].

Chapter 6

Generalised phase kick-back: the physical structure of algorithms

As discussed the start of the last chapter, our investigations into computation within the generalised probabilistic theory framework thus far, comprising chapters 2, 3, and 4, have taken a high-level approach, using the language of complexity classes to derive general bounds on the power of computation. However, much of quantum computing is concerned not so much with this high-level view, but instead with the construction of concrete algorithms to solve specific problems. A deeper understanding of the general structure of computational algorithms in this framework has so far remained illusive. Here we take a low-level algorithmic view and ask which physical principles are required to allow for some of the common machinery of quantum computation in this context.

In this chapter we show that the principles of causality, purification, purity preservation, and strong symmetry (principles 2 to 5 from chapter 1) are sufficient for the existence of *reversible controlled transformations* and a generalised *phase kick-back mechanism*. In the quantum case, the phase kick-back mechanism [135] plays a vital role in almost all algorithms—notably the Deutsch-Jozsa algorithm [134], Grover's search algorithm [149] and Simon's algorithm [150]—whilst reversible controlled transformations are central components of most well-studied universal gate sets, and are fundamental for the definition of computational oracles.

We use these results to investigate the relationship between interference behaviour and computational power in theories satisfying the above principles. Recall from chapter 5 that Sorkin [124, 125] has introduced a hierarchy of mathematically conceivable *higher-order* interference behaviours, and shown that quantum theory is limited to having only second-order interference. This second-order interference between quantum computational paths is a resource for post-classical computation [84, 1]. It therefore seems prudent to investigate whether the connection between post-classical interference (i.e. $n > 1$ in Sorkin's hierarchy) and post-classical computation holds in general.

In quantum theory there is an intimate connection between phase transformations—such as those used in the kick-back mechanism—and interference. Motivated by this, we introduce a framework that relates higher-order interference to *phase transformations* in operationally-defined theories. We show that the generalised phase kick-back mechanism allows one to access any 'higher-order phase' in a controlled manner. Using this, we show that the existence of non-trivial interference behaviour allows for the solution of problems intractable on a classical computer, demonstrating that post-classical interference is a general resource for post-classical computation. In proving the above, we connect higher-order phases to the existence of new particle types which exhibit both qualitatively and quantitatively different behaviour to fermions, bosons, and anyons; potentially providing a novel experimental test of higher-order interference.

Lastly, we use the existence of reversible controlled transformations to provide a rigorous definition of a computational oracles, solving a problem that was posed in chapter 1. Moreover, we show that in theories satisfying the above principles there is a general solution to the subroutine problem, which was first discussed in the quantum case by [44]. That is, we use the existence of reversible controlled transformations to show that $\mathbf{BGP}^{\mathbf{BGP}} = \mathbf{BGP}$.

6.1 Higher-order interference via phase transformations

6.1.1 A quantum example

Perhaps the cleanest example of interference in quantum theory is exhibited by the Mach-Zehnder interferometer, illustrated in Fig. 6.1. There are three parts to this:

1. Prepare a state as a superposition of paths:

$$|s) = |+\rangle\langle+| := \rho_+$$

2. Apply a 'phase transformation':

$$P_{\Delta\phi}|s) = R_z^{\Delta\phi}\rho_+ R_z^{\Delta\phi\dagger},$$

with $R_z^{\Delta\phi}$ a rotation by $\Delta\phi$ about the z axis of the Bloch ball.

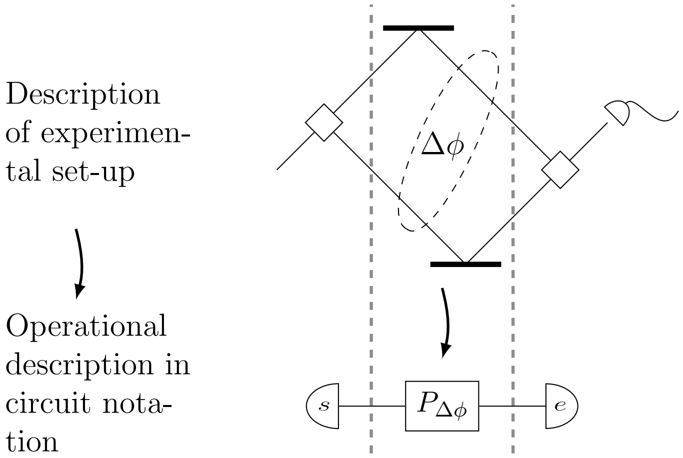

Description of experimental set-up
$\Delta\phi$
Operational description in circuit notation
s
$P_{\Delta\phi}$
e

Fig. 6.1: Relation between the physical set-up of Mach-Zender interferometer and the operational description, depicted in circuit notation

3. Measure in a superposition of paths:

$$(e|P_{\Delta\phi}|s) = \mathrm{Tr}\left(\rho_+ R_z^{\Delta\phi}\rho_+ R_z^{\Delta\phi\dagger}\right) = \cos^2\left(\frac{\Delta\phi}{2}\right).$$

The observed interference pattern is therefore a map from the group of 'phase transformations', parametrised by $\Delta\phi$, to the unit interval (i.e. probabilities),

$$P_{\Delta\phi} \mapsto \cos^2\left(\frac{\Delta\phi}{2}\right). \tag{6.1}$$

The existence of interference in quantum theory is encapsulated in the statement: "the interference pattern observed for a particular superposition measurement cannot be reproduced by the statistics generated by 'which path' measurements". In the above example this translates to:

$$\cos^2\left(\frac{\Delta\phi}{2}\right) \neq \sum_{i=0}^{1} q_i \mathrm{Tr}\left(|i\rangle\langle i| R_z^{\Delta\phi}\rho_+ R_z^{\Delta\phi\dagger}\right), \tag{6.2}$$

where q_i is an arbitrary constant. Equation 6.2 is to be interpreted as an inequality of the functions defined on the right and left hand side. That is, these functions do not coincide on all phase transformations. This follows from the fact that:

$$R_z^{\Delta\phi\dagger}|i\rangle\langle i|R_z^{\Delta\phi} = |i\rangle\langle i|, \quad \forall i \in \{0,1\}. \tag{6.3}$$

That is, the left hand side of equation 6.2 depends on $\Delta\phi$ whilst the right hand side does not.

6.1.2 Operational theories

The quantum example from section 6.1.1 illustrates the key components necessary to discuss interference:

(i) a notion of 'path',

(ii) a notion of 'superposition of paths',

(iii) transformations that leave the statistics of 'which path' measurements invariant, i.e. 'phase transformations',

(iv) a notion of 'interference pattern', i.e. a way of associating phase transformations with probabilities.

These points will now be discussed in the context of arbitrary operationally-defined theories. We then use this framework to link higher-order interference and phase transformations. Our approach is similar in spirit to that of Garner et al. [130], with the caveat that they have not considered higher-order interference.

—**(i) Paths:** A path is defined by a state and effect pair, where we view the state as 'preparing a state which belongs to the path' and the effect as 'measuring whether the state belongs to the path' and so we demand the probability of the state-effect pair to be one.

Definition 6.1. *Paths, p:*

$$p := (|s), (e|) \ s.t. \ (e|s) = 1.$$

In our quantum example, the paths were $p_0 = (|0\rangle\langle 0|, |0\rangle\langle 0|)$ and $p_1 = (|1\rangle\langle 1|, |1\rangle\langle 1|)$.

Paths are disjoint if the state defining one path has zero probability of belonging to the other, and vice versa.

Definition 6.2. *Disjoint paths, $p_1 \perp p_2$:*

$$p_1 \perp p_2 \iff (e_i|s_j) = \delta_{ij}.$$

An n-path experiment is defined by n mutually disjoint paths such that the set consisting of the effects from each path forms a measurement.

Definition 6.3. *n-path experiment, $\mathbb{P}$:*

$$\mathbb{P} := \{p_i\} \ s.t. \ p_i \perp p_j \ \forall i \neq j, \text{ and } \sum_i (e_i| = (u|.$$

Where $(u|$ in the above definition is the unique deterministic effect. In the quantum case, an n-path experiment would correspond to a multi-arm interferometer.

—**(ii) Superposition of paths:** A superposition of paths will be defined relative to some n-path experiment $\mathbb{P}$ via the notion of *support.* We say that a state (or effect) has support on a path if the effect (or state) associated to that path gives a non-zero probability.

Definition 6.4. *Support of a state or effect, $Supp[|s)]$ or $Supp[(e|]$:*

$$\begin{aligned} Supp[|s)] &:= \{p_i \in \mathbb{P} \mid (e_i|s) \neq 0\}, \\ Supp[(e|] &:= \{p_i \in \mathbb{P} \mid (e|s_i) \neq 0\}. \end{aligned}$$

If the support of a state consists of more than one path this does not guarantee that it is a superposition of paths—it could equally well be a classical mixture of paths. A superposition state must therefore lie outside the convex hull of the states which have support only on a single path. In our quantum example, the state $|+\rangle\langle+|$—introduced in point 1 of section 6.1.1—was a superposition of paths.

We can define set of states (or effects) with support on some subset of paths $I \subseteq \mathbb{P}$ as:

$$\begin{aligned} \Omega_I &:= \{|s) \in \Omega \mid Supp[|s)] = I\}, \\ \mathcal{E}_I &:= \{(e| \in \mathcal{E} \mid Supp[(e|] = I\}. \end{aligned}$$

—**(iii) Phase transformations:** A phase transformation—relative to some $\mathbb{P}$—is any transformation that leaves the statistics of 'which path' measurements invariant.

Definition 6.5. *Phase group, $\mathcal{P}$:*

$$\mathcal{P} := \{T \in \mathcal{R} \mid (e_i|T = (e_i|, \ \forall i \in \mathbb{P}\}$$

In the quantum example, the phase transformation was the rotation $R_z^{\Delta\phi}$ introduced in point 2 of section 6.1.1.

—**(iv) Interference patterns:** We now generalise the quantum interference pattern of equation 6.1 to arbitrary operational theories.

Definition 6.6. *Interference pattern, $\mathcal{C}_{s,e}$:*

$$\mathcal{C}_{s,e} : \ \mathcal{P} \to [0,1] \ :: \ T \mapsto (e|T|s)$$

Given this definition, the existence of quantum interference, as represented in equation 6.2, translates into the existence of $(e| \in \mathcal{E}_{\{0,1\}}$—that is, an effect with support on path 0 and path 1—and $|s) \in \Omega_{\{0,1\}}$ such that

$$C_{s,e} \neq \sum_{i=0}^{1} \mathcal{C}_{s,e_i} \tag{6.4}$$

for all possible choices of $(e_i| \in \mathcal{E}_{\{i\}}$ (including sub-normalised effects, which is the analogue of the q_i's in equation 6.2). In other words, there is some choice of superposition state and effect such that their interference pattern cannot be reproduced by the statistics generated by any set of effects with support on a single path.

It follows that the existence of a non-trivial phase group implies the existence of interference in a general theory. Indeed, the left hand side of equation 6.4 depends on the phase group element, whilst the right hand side does not—the analogue of equation 6.3 from the quantum example. We now use our framework to discuss *higher-order* interference.

6.1.3 Higher-order interference and phase

As we saw in chapter 5, other approaches to defining higher-order interference in operational theories (that of Barnum et al. in [72] for example) require a theory to have additional structure if the definition is to make sense. Such extra structure results in transformations that represent the action of leaving open some subset of paths I, whilst blocking the others. In this case one can define $(e_I| = (e|P_I$, P_I being the transformations that leave subset of paths I open, giving a specific set of effects with which one can check for the existence of higher-order interference. However, as we saw explicitly in chapter 5 for Quartic Quantum Theory, arbitrary theories do not necessarily have such structure. Hence, when checking for the existence of higher-order interference in an arbitrary theory, one must consider all possible choices $(e_I|$ with the correct support. Otherwise, as was the case in Quartic Quantum Theory, one can choose a specific set of $(e_I|$ to give the artificial appearance of higher-order interference.

The following adaptation of Sorkin's original definition of higher-order interference [124] to our framework of phase transformations results in the following. This definition is applicable to any generalised probabilistic theory and does not require further structure.

Definition 6.7. *The existence of nth-order interference in an n-path experiment corresponds to the existence of an effect $|e)$ and a state $(s|$, such that*

$$\mathcal{C}_{s,e}(T) \neq \sum_{I \subset \mathcal{P}} (-1)^{n-|I|+1} \mathcal{C}_{s,e_I}(T), \tag{6.5}$$

$\forall (e_I| \in \mathcal{E}_I$.

The introduction of the '$\forall$' statement compared to the original definition is due to the ambiguity in choosing which effect corresponds to blocking some subset of paths. Henceforth, the explicit dependence on T in the above equation will be suppressed, as $\mathcal{C}_{se}$ has already been defined as a function from the phase group to probabilities. As in equation 6.2, equation 6.5 is to be interpreted as an inequality of the functions defined on the right and left.

For example, the existence of second-order interference implies that there exists $|s)$ and $(e|$ such that

$$\mathcal{C}_{s,e} \neq \mathcal{C}_{s,e_{\{0\}}} + \mathcal{C}_{s,e_{\{1\}}}, \ \forall \ |e_{\{i\}}) \in \mathcal{E}_i.$$

Whilst the existence of third-order interference corresponds to the existence of, $|s)$ and $(e|$ such that

$$\mathcal{C}_{s,e} \neq \mathcal{C}_{s,e_{\{0,1\}}} + \mathcal{C}_{s,e_{\{1,2\}}} + \mathcal{C}_{s,e_{\{2,0\}}} - \mathcal{C}_{s,e_{\{0\}}} - \mathcal{C}_{s,e_{\{1\}}} - \mathcal{C}_{s,e_{\{2\}}}, \ \forall \ |e_I) \in \mathcal{E}_I.$$

The above definition of higher-order interference produces the expected results when applied to classical or quantum theory, as it should. To illustrate this, we now show the existence of second-order interference in quantum theory via this new definition. The calculation showing the lack of third-order interference in quantum theory—relative to this new definition—is provided in appendix D. Define our paths by $p_i := (|i\rangle\langle i|, |i\rangle\langle i|)$, then choose $|s) = |+\rangle\langle +| = (e|$. Then $(e_{\{i\}}| \in \{r_i|i\rangle\langle i|\}$ where r_i is an arbitrary positive real number. The phase group is given by $\mathcal{P} := \{e^{i\theta_0}|0\rangle\langle 0| + e^{i\theta_1}|1\rangle\langle 1|\}$. It is then simple to show that,

$$\mathcal{C}_{s,e}(T) = \cos^2\left(\frac{\theta_0 - \theta_1}{2}\right),$$

whilst,

$$\mathcal{C}_{s,e_{\{0\}}}(T) + \mathcal{C}_{s,e_{\{1\}}}(T) = \frac{r_0 + r_1}{\sqrt{2}}.$$

It is then simple to see that, as functions of θ_i,

$$\cos^2\left(\frac{\theta_0 - \theta_1}{2}\right) \neq \frac{r_0 + r_1}{\sqrt{2}},$$

for any choice of r_i, i.e. $(e_{\{i\}}|$. Therefore—by the above definition—quantum theory has second-order interference, as we would expect. In appendix D we provide a more in-depth comparison between Sorkin's original definition and the one presented above.

Motivated by equation 6.5, we wish to determine if particular phase transformations give rise to higher-order interference. The defining feature of phase transformations is that they leave the statistics of effects with support on single paths—that is, effects in $\bigcup_i \mathcal{E}_{\{i\}}$—invariant. The natural generalisation of this is to consider transformations that not only leave the statistics of effects on single paths invariant, but also superposition effects. This motivates the following definitions.

Definition 6.8. *A transformation T is n-*undetectable *if:* $(e|T = (e|,\ \forall (e| \in \bigcup_{I:|I|\leq n} \mathcal{E}_I$.

Together with its natural converse.

Definition 6.9. *A transformation T is m-*detectable *if there exists $(e| \in \bigcup_{I:|I|\leq m} \mathcal{E}_I$, such that $(e|T \neq (e|$.*

We can now link higher-order interference to certain types of phase transformations, which we call *higher-order phases.*

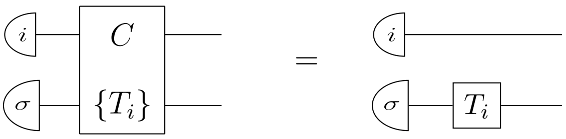
Theorem 6.10. *A transformation T that is n detectable and $n-1$ undetectable implies the existence of nth-order interference.*

Proof. Choose $|s)$ and $(e|$ such that T is detected. It is then clear that the left hand side of equation 6.5 is dependent on T, whilst—due to undetectability—the right hand side is not. They are thus distinct functions. □

In our quantum example, the phase transformation was 2-detectable, but 1-undetectable.

6.2 Controlled transformations and a generalised phase kick-back

Definition 6.11. *Given a set of pure and perfectly distinguishable states $\{|i)\}$, and a set of transformations $\{T_i\}$, define a controlled transformation $C\{T_i\}$ as:*



$\forall i, |\sigma)$ (6.6)

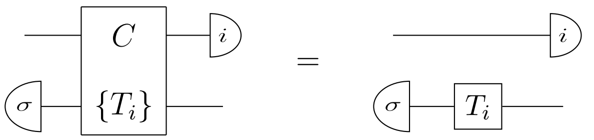
The top system and lower systems are referred to as the control *and* target *respectively.*

Note that classical controlled transformations—where the control is measured and conditioned on the outcome a transformation is applied to the target—exist in any causal theory [15] with sufficient distinguishable states, such as Boxworld, or indeed any of the polygon theories briefly mentioned in chapter 1. However, such transformations are in general not reversible and do not offer an advantage over classical computation [11]. Moreover, the existence of reversible controlled transformations appears to be a rare property of operational theories [11]. The following states that in theories satisfying physical principles 2 to 5, there exist reversible controlled transformations. The proof is in contained in section 6.4.

Theorem 6.12. *In any theory satisfying i) causality, ii) purification, iii) purity preservation, iv) strong symmetry, and in which there exists a set of n pure and perfectly distinguishable states, there exists a* reversible *controlled transformation for any collection of n reversible transformations $\{T_i\}_{i=1}^n$.*

Moreover, the following lemma states that any controlled transformation in such theories 'preserves superpositions'. Where 'superposition' is meant in the sense of section 6.1.2 part (ii) and 'preserves superposition' means that the probability of detecting the system in each path of the superposition is preserved by the transformation. See section 6.5 for the proof.

Lemma 6.13. *Superpositions are preserved on the control input:*

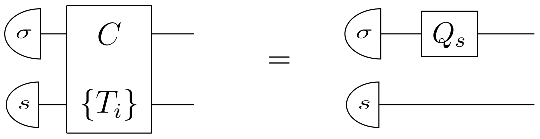


$\forall i,\ |\sigma)$ (6.7)

where $\{(i|\}$ is the measurement that perfectly distinguishes the control states $\{|i)\}$ [1].

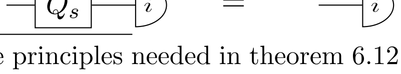
Every controlled transformation in quantum theory has a *phase kick-back* mechanism [1, 135]. Such mechanisms form a vital component of most quantum algorithms [135]. We now show the existence of a *generalised* phase kick-back mechanism in any theory satisfying our assumptions. The proof is provided in section 6.6.

Lemma 6.14. *Given an $|s)$ such that $T_i|s) = |s)$, $\forall i$, there exists a reversible transformation Q_s such that*



$\forall|\sigma)$ (6.8)

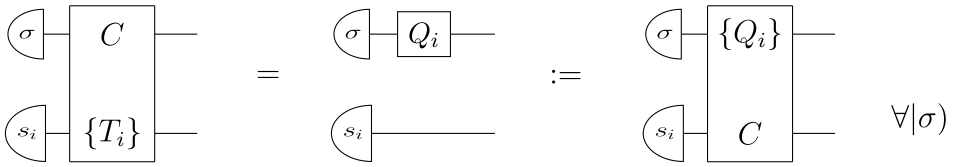
Moreover, Q_s is phase transformation:



$\forall i$

[1] In theories satisfying the principles needed in theorem 6.12, the measurement $\{(i|\}$ is unique up to normalisation, see section 6.3.1.

In quantum theory, it is possible to achieve any phase transformation via a kick-back mechanism. However, lemma 6.14 only implies the existence of at least one phase that can be 'kicked-back'. We now show that all phases arise via the generalised mechanism. Consider the set of pure and perfectly distinguishable states $\{|s_i)\}$ and let $\{T_i\}$ be elements of their phase group, i.e. $T_i|s_j) = |s_j)$, $\forall i, j$. Construct the controlled transformation $C\{T_i\}$. The designation of control and target for $C\{T_i\}$ is symmetric:



Thus any $C\{T_i\}$ is control-target symmetric if the T_i are elements of a phase group. The transformations on the target are given by the kicked-back phases, $\{Q_i\}$. Given an arbitrary W_i, construct the transformation $\{W_i\}C$ and note that via control-target symmetry it is equivalent to $C\{G_i\}$, for some $\{G_i\}$. The controlled transformation $C\{G_i\}$ thus gives rise to the kicked-back phase W_i and we have:

Theorem 6.15. *Every phase transformation can arise via a generalised phase kick-back mechanism*

6.2.1 Particle exchange experiments

Dahlsten et al. [136] have shown that there is a close connection between particle exchange statistics and the phase group in operational theories. We use the framework and results presented in this paper to expand upon and formalise these connections. Motivated by the quantum case, place a pair of indistinguishable particles in superposition by inputting them to an interferometer, as shown in the Fig. 6.2. On the upper path the two particles are swapped using some operation 'S', whilst on the lower path they are left invariant—that is, the identity operation $\mathbb{I}$ is applied. The entire physical set-up is described by a bipartite state, one partition of which corresponds to the state of the particles, $|s) \in \Omega_{P'cles}$, and the other to the 'which path' information embodied in the interferometer, $|s') \in \Omega_{Path}$. The entire scenario thus takes place in the state space $\Omega_{Path} \otimes \Omega_{P'cle}$. In the quantum case, the phase transformation generated by this procedure corresponds to the type of indistinguishable particle employed in the experiment.

The whole experiment can be described via a controlled transformation, with path information as the control and particle state as the target. Via theorem 6.12, such

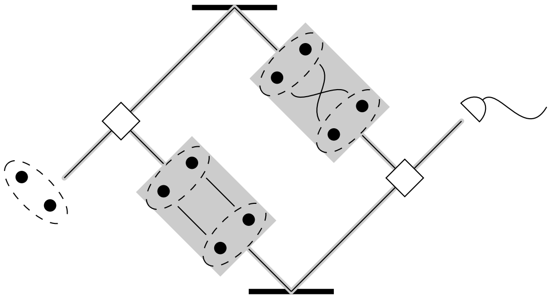
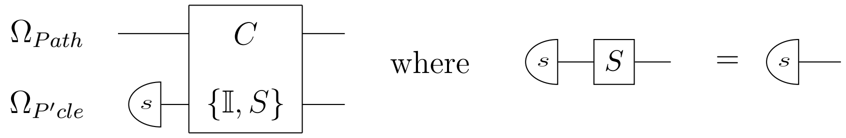


Fig. 6.2: Schematic illustration showing a particle exchange experiment involving two particles.

an experiment exists in theories satisfying our assumptions. Applying operation S to the particle state corresponds to swapping a pair of indistinguishable particles and so must leave the statistics of any measurement invariant. Therefore $S|s) = |s)$, where $|s)$ is the initial particle state.



Theorem 6.14 tells us that the above diagram corresponds to a kicked-back phase on the control system, as in equation 6.8. Thus, to every particle type, there exists a corresponding phase transformation, which was the connection discussed in [136]. But, in an arbitrary theory, the converse is not necessarily true. The quantum phase group is $U(1)$ and, fixing its representation to be $\{e^{i\theta}\}$, bosons kick-back the transformation corresponding to $\theta = 0$, fermions $\theta = \pi$ and anyons any arbitrary θ. Thus, to every particle type in quantum theory there is an associated phase, and vice versa.

To generalise this to theories satisfying our three assumptions, we must connect the operational description of these theories to the more physical notion of particles. Towards this end, we make the following two assumptions:

1. Every operational state $|s)$ corresponds to the state of some collection of indistinguishable particles,

2. Every transformation that leaves the operational state $|s)$ invariant corresponds to a (possibly trivial) permutation of the collection of indistinguishable particles.

Given the above, theorem 6.15 tells us that to every phase transformation there exists a corresponding particle type. Therefore, to each higher-order phase—described in

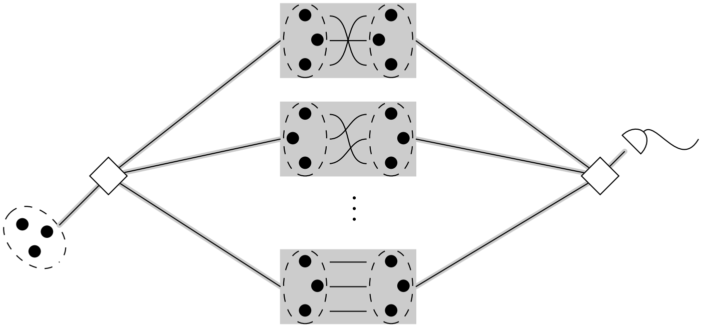


Fig. 6.3: Schematic illustration showing a particle exchange experiment involving mulitple particles.

theorem 6.10—there is associated a particle type that should be observable through a generalisation of the above experiment.

Consider $|s)$, which corresponds to the state of some collection of indistinguishable particles, and a permutation operation π which leaves $|s)$ invariant. Note that, for a given permutation, there may be multiple topologically distinct ways of performing it, particularly in two dimensions or topologically non-trivial spaces. Now consider the n-path experiment, illustrated in Fig. 6.3, where on each path some distinct permutation operation $\pi_i,\ i = 1, \ldots, n$, takes place. This can be described by a controlled transformation $C\{\pi_i\}$. Given the above two assumptions, any nth-order phase—if they exist in the operational theory—can be kicked-back by such an experiment.

Recall that nth-order phases are n-detectable, but $n-1$-undetectable. That is, the action of such phases cannot be detected by any effect with support on less than n paths. Which is in stark contrast to quantum, or 2nd-order, phases, which can always be detected by an effect with support on two paths. Thus, permutations of particles, whose type all correspond to an nth-order phase, can only be detected by recombining *all* paths in an n-path experiment. In some sense then, nth-order phases encode holistic information about all paths in an n-path experiment.

6.2.2 Computational oracles

Oracles play a vital role in quantum computing, forming the basis of most known speed-ups over classical computation [1]. As we saw in section 2.4 from chapter 2, defining a general notion of oracle—that reduces to the standard notion in the quantum case—in operationally-defined theories appears to be a difficult problem. A particular example of a quantum oracle is the following controlled unitary:

$$U_f = |0\rangle\langle 0| \otimes Z^{f(0)} + |1\rangle\langle 1| \otimes Z^{f(1)}, \tag{6.9}$$

with Z a Pauli matrix, $f : \{0,1\} \to \{0,1\}$ a function encoding some decision problem and $Z^0 := \mathbb{I}$. The quantum phase kick-back for U_f amounts to

$$U_f = \mathbb{I} \otimes |0\rangle\langle 0| + Z^{f(0)\oplus f(1)} \otimes |1\rangle\langle 1|. \tag{6.10}$$

One can see that inputting $|+\rangle|1\rangle$ and measuring the first qubit in the $\{|+\rangle, |-\rangle\}$ basis reveals the value of $f(0) \oplus f(1)$ in a single query of the oracle—a feat impossible on a classical computer [1].

The results of theorem 6.12 provide a way to define computational oracles in any theory satisfying our assumptions, answering a question which was posed in chapter 2. An oracle in such theories corresponds to a reversible controlled transformation[2] where the set of transformations $\{T_{i,f(i)}\}$ being controlled depend on a function $f : \{i\} \to \{0,1\}$ encoding a decision problem of interest. As the transformations $T_{i,f(i)}$ depend on the value of $f(i)$, so does the controlled transformation and the kicked-back phase. That is, in theories with a non-trivial phase group, the phase kick-back of an oracle encodes information about the value $f(i)$ for all i. In such theories, there is thus a non-zero probability of extracting such global information. Non-trivial interference behaviour can thus be seen as a general resource for non-classical computation.

In the quantum case, there is a limit to how much global information one can obtain in a single oracle query. In the situation where $f : \{0,\ldots,n-1\} \to \{0,1\}$ a quantum oracle can extract the value of $f(i) \oplus f(j)$, for some i, j, in a single query without error [1]. Can theories with higher-order interference reliably extract more global information about f—without error—in a single query? In chapter 5 we saw that the theory of Density Cubes offers a slight advantage over quantum theory in a certain computational task. The results of section 6.2.1 appear to suggest that nth-order phases encode information about all paths in an n-path experiment, as opposed to 2nd-order, or quantum, phases which only encode information about at most two paths. This fact, in conjunction with theorem 6.15, suggests the possibility that theories with higher-order interference may be able to solve problems intractable even on a quantum computer. This possibility is investigated from the point of view of the search problem in the next chapter.

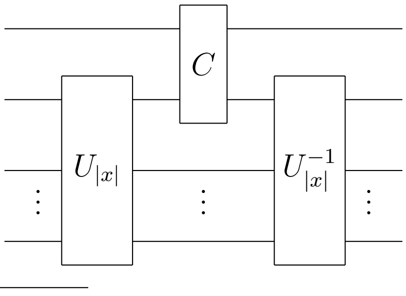
Now that we have a rigorous definition of an oracle, a natural question is whether **BGP** is closed under *subroutines*. Another way to pose this question is to ask whether $\mathbf{BGP}^{\mathbf{BGP}} = \mathbf{BGP}$ for $\mathbf{G}$ satisfying causality, purification, purity preservation, and

[2] There could be many distinct transformations that have the same behaviour on a set of control states. As long as one fixes which transformation corresponds to the oracle, this is not a problem.

strong symmetry? That is, does having access to an oracle[3] for a particular decision problem which can be efficiently solved in that theory provide any more computational power than just using the efficient algorithm? A potential issue arises when one compares the performance of the oracle implementation to that of the efficient algorithm when both are used as subroutines in another computational algorithm. Here, an algorithm consists of a poly-size, uniform circuit family where for simplicity, as in chapter 4, we fix the acceptance function $a(.)$ to only depend on a final measurement applied to the first output system. Moreover, to move closer to the quantum accepetence condition, this final measurement will always correspond to $\{(0|, (1|\}$, where each $(i|$ is a pure effect, and outcome 0 corresponds to the accept. As we have seen above, oracles can be queried in superposition, but one does not usually query an algorithm for a particular decision problem in superposition; one merely prepares the state corresponding to a particular bit string and uses the algorithm to determine whether or not that bit sting is in the language in question. Does every **BGP** algorithm for a decision problem admit a subroutine having the characteristics of an oracle for that decision problem? Such a was proved in the quantum case by Bennett et al. in [44]. The following theorem shows that it is also true for theories satisfying our 4 principles.

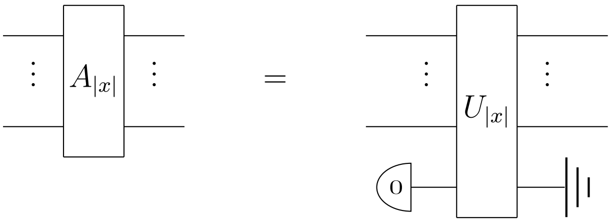
Theorem 6.16. *Consider a theory* $\boldsymbol{G}$ *which satisfies causality, purification, purity preservation, and strong symmetry. Given an algorithm* $\{A_{|x|}\}$ *for a decision problem in* ***BGP****, one can always construct a circuit family* $\{C_{|x|}\}$*, consisting of reversible transformations from* $\boldsymbol{G}$*, which, with high probability, functions as an oracle for that particular decision problem. Hence,* $\mathbf{BGP}^{\mathbf{BGP}} = \mathbf{BGP}$.

It was shown in [15] that the purification principle implies the ability to dilate any transformation to a reversible one. We use this fact in the construction of the circuit $\{C_{|x|}\}$. Our construction is equivalent to the one employed in the quantum case [44]. Each $C_{|x|}$ is corresponds to

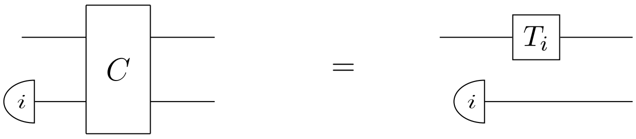


[3] represented as some function which can be queried using said oracle

where U is the reversible transformation which dilates[4] the **BGP** algorithm $A_{|x|}$

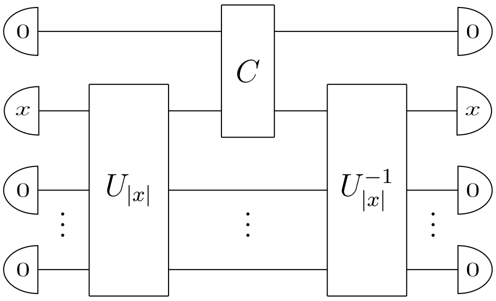


and C is a reversible controlled transformation with the lower system as the control



with $|i) \in \{|0), |1)\}$, $T_0 = \mathbb{I}$, and where T_1 acts as $T_1|i) = |i \oplus 1)$.

To prove theorem 6.16, we need to show that the probability corresponding to the following closed circuit



is greater than or equal to $1 - 2^{-q(|x|)}$, for some polynomial $q(|x|)$, when x is in the required language, as this entails that $C_{|x|}$ functions as an oracle with high probability. The proof of theorem 6.16 is presented in section 6.7 and is essentially the analogue of the original quantum proof, but generalised and adapted to our setting.

6.3 General results following from physical principles 2 to 5

To prove our results we will need the following consequences of principles 2 to 5.

[4] Here we assume for simplicity that the circuit family $\{U_{|x|}\}$, with $U_{|x|}$ a reversible transformation which dilates $A_{|x|}$ for each $|x|$, consists of poly-size uniform circuits.

6.3.1 Self-duality and uniqueness of distinguishing measurements

Principles 2 to 4, together with the existence of (at least two) pure and perfectly distinguishable states, imply the following result (see [144] for a proof): for any given state $|s)$, there exists a natural number n and a set of pure and perfectly distinguishable states $\{|a^i)\}_{i=1}^n$ such that $|s) = \sum_i p_i|a_i)$ where $0 \leq p_i \leq 1$, $\forall i$ and $\sum_i p_i = 1$.

This result, together with principle 5, implies the existence of a "self-dualising" [49, 72] inner product $\langle \cdot, \cdot \rangle$, briefly discussed in chapter 4. That is, to every pure state $|s)$, there is associated a unique pure effect $(e^s|$, satisfying $(e^s|s) = 1$, such that: $(e^s|(\cdot) = \langle |s), \cdot \rangle$. Moreover, the conjunction of causality, purification, purity preservation, and strong symmetry with the existence of a set of pure and perfectly distinguishable states $\{|i)\}$ implies the existence [49, 72] of a unique measurement $\{(j|\}$ such that

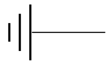
$$(i|j) = \delta_{ij}.$$

That is, if there exists an effect $(r|$ satisfying $(r|i) = 1$ it must be the case that $(r| = (i|$. Additionally, if there is a set $\{(e_j|\}$ such that $(e_j|i) = \alpha_j \delta_{ij}$ for all j then,

$$(e_j| = \alpha_j (j|.$$

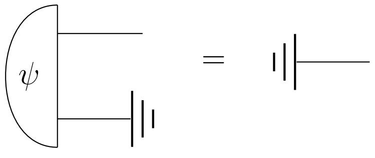
6.3.2 Existence of a completely mixed state

Recall from chapter 1 that purification implies the existence of a unique *completely mixed* state $|c)$, defined by $T|c) = |c)$ for all reversible transformations T. Any state is a 'refinement' of this state. See chapter 1 and [15, 16] for a more in-depth discussion. The completely mixed state will be represented diagrammatically as:

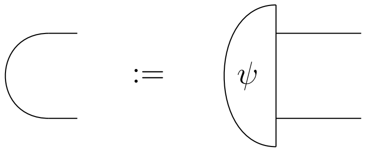


6.3.3 Purification of the completely mixed state is dynamically faithful

The purification principle implies the existence of a pure state $|\psi)$ which purifies the completely mixed state:

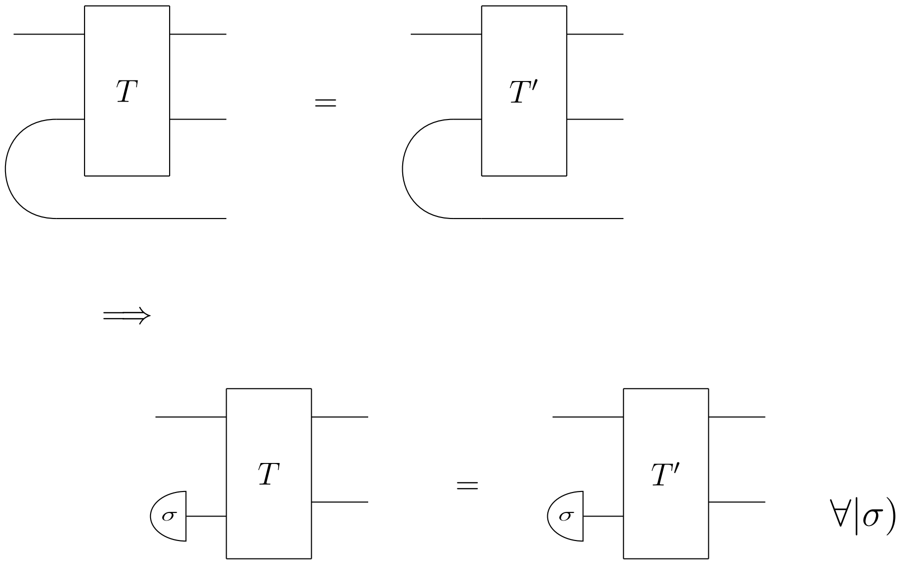


This is unique up to reversible transformation on the purifying system. In quantum theory, an example of such a $|\psi)$ would be the maximally entangled state $|\psi\rangle\langle\psi|$ with $|\psi\rangle = \sum_i |i\rangle|i\rangle/\sqrt{N}$. For ease and simplicity of notation, we fix a diagrammatic representation of a particular choice of this purification as follows:



Note that, despite the seeming symmetry in the above diagrammatic representation, the above pure state state is not required to be symmetric, i.e. it is not required that applying the unique deterministic effect to the top system results in a completely mixed state on the bottom system. The above representation is only chosen to make the diagrammatic proofs contained in the next few sections easier to follow.

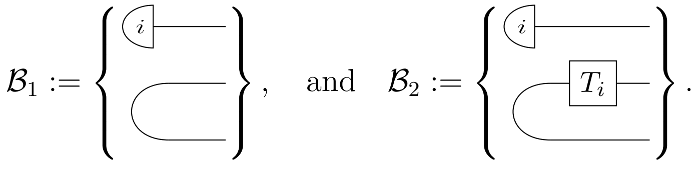
Purifications of the completely mixed state are called *dynamically faithful* states [15, 16] and satisfy the following important condition [15, 16], known as *dynamic faithfulness*:



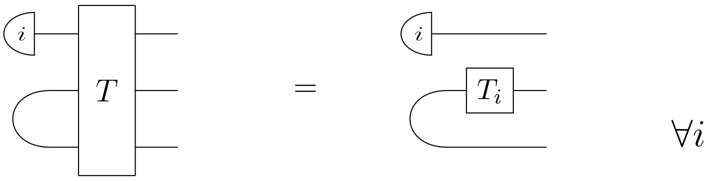
In quantum theory, the above implication is known as Choi's theorem [15].

6.4 Proof of theorem 6.12

Recalling that purity preservation implies that the composite of pure states is pure, we can define two sets of pure and perfectly distinguishable states:

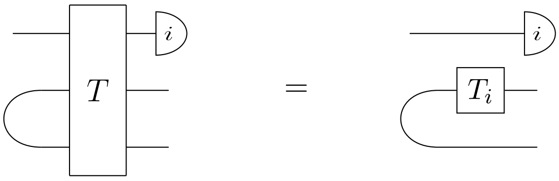


Strong symmetry implies that there exists a reversible transformation between these two sets, $T : \mathcal{B}_1 \to \mathcal{B}_2$.

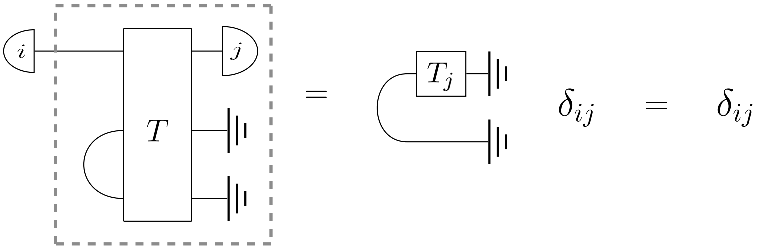


This result, together with the existence of dynamically faithful states, will be used to show the existence of a reversible controlled transformation $C\{T_i\}$ for an arbitrary set of reversible transformations $\{T_i\}$. We break the proofs into lemmas, which we now present.

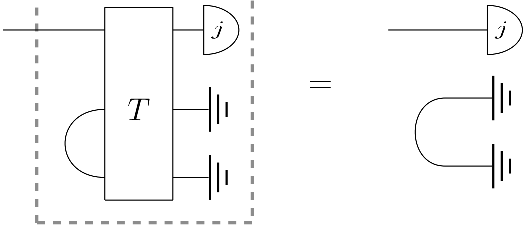
Lemma 6.17. *'Superposition preservation'*



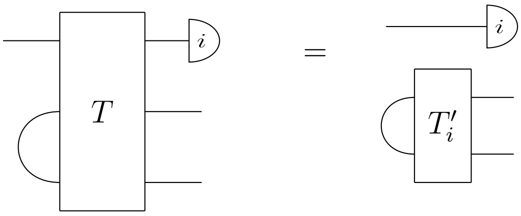
Proof. Firstly we prove a weaker condition which is superposition preservation for the following,



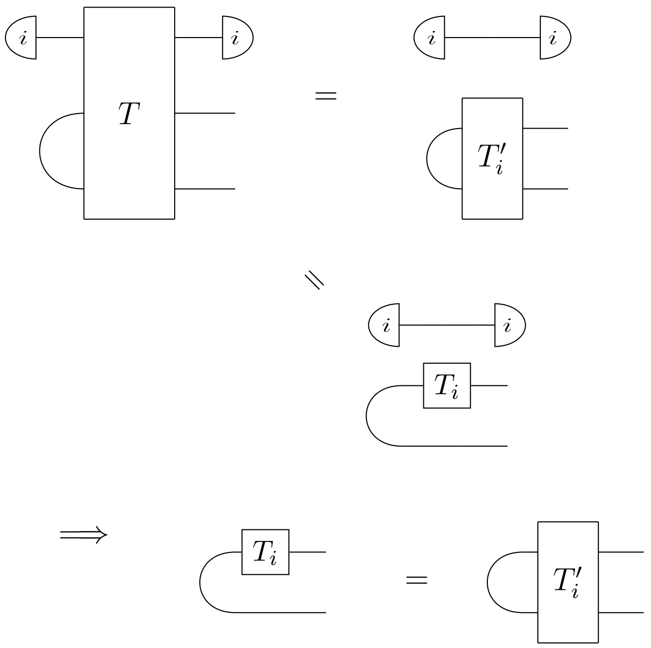
Strong symmetry $\implies$



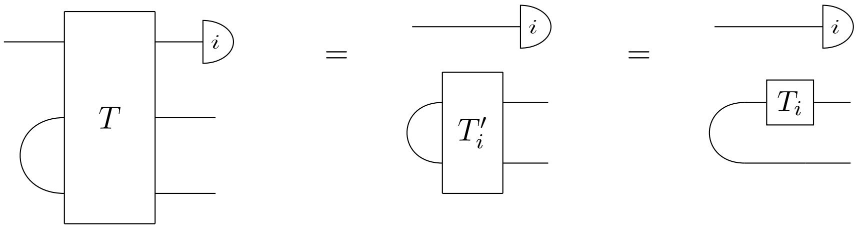
the implication follows from the uniqueness of the maximally distinguishing measurement and the fact that the dynamically faithful state is normalised. Then purification implies,



Now consider,

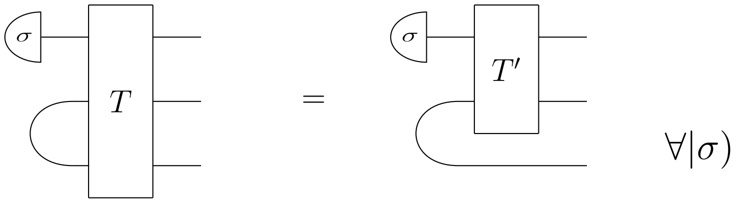


where the above follows the fact that $(i|i) = 1$. This, in conjunction with the previous results, gives:

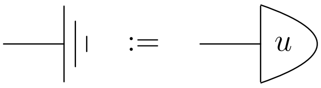


□

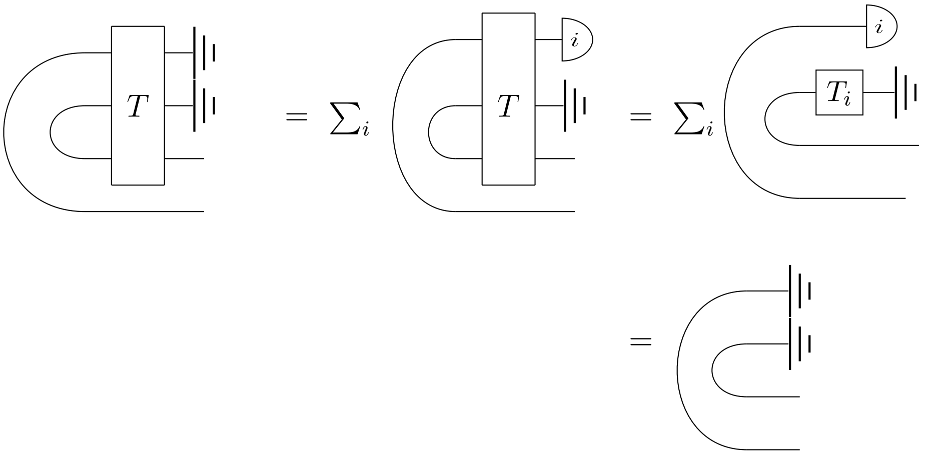
Lemma 6.18. *There exists a reversible transformation T', such that*



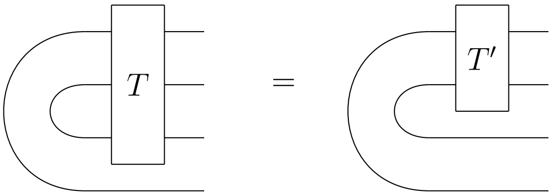
Proof. Recall from chapter 1 that causality implies $\sum_i (i| = (u|$, where $\{(i|\}$ is the measurement which perfectly distinguishes the control states $\{|i)\}$, and $(u|$ the unique deterministic effect



We thus have



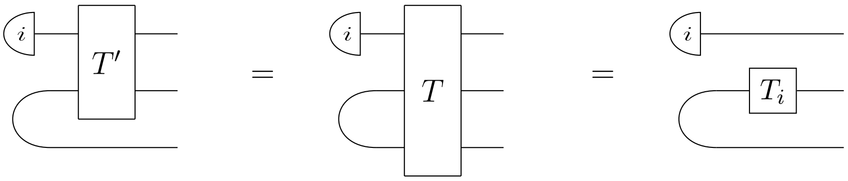
Purification $\Longrightarrow$



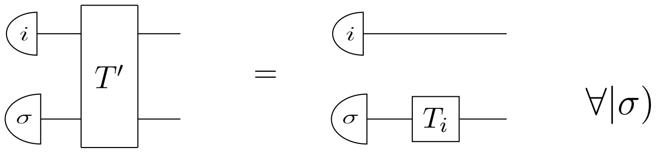
Dynamic faithfulness then gives the result. □

Theorem 6.19. *T' is a reversible controlled transformation, $T' = C\{T_i\}$.*

Proof.



Dynamically faithful state $\Longrightarrow$

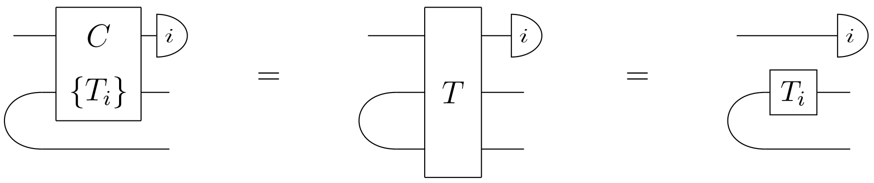


which is the defining characteristic of $C\{T_i\}$. □

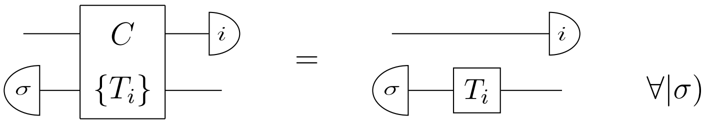
6.5 Proof of lemma 6.13

Lemma 6.17 already provides some notion of superposition preservation, we can use the other results above to extend this.

Proof.



The first equality uses theorem 6.19 and the second lemma 6.17. Using the result of dynamically faithful states then implies,

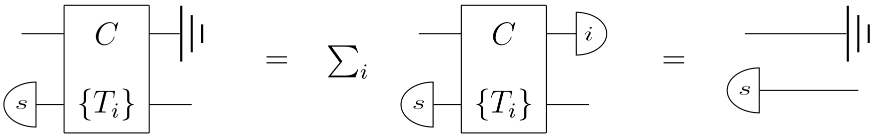


□

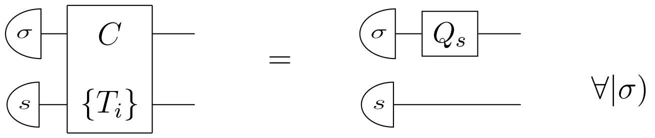
This actually only proves the existence of a controlled transformation that preserves superpositions, it is simple to show that it must be true for all controlled transformations using an argument analogous to lemma 6.17.

6.6 Proof of lemma 6.14

Proof.

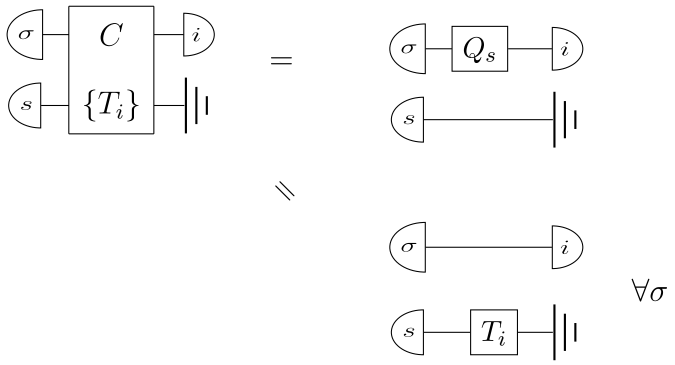


The first equality follows from causality and the second from equation 6.7 and the definition of $|s)$. Applying this transformation to the top half of the dynamically faithful state and using purification results in the existence of a reversible Q_s such that:

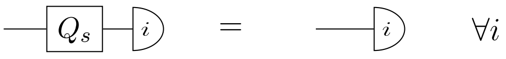


Note that Q_s depends on both the controlled transformation and the joint eigenstate

$|s)$. Note that:



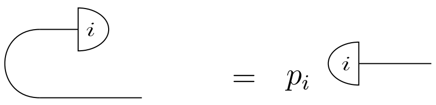
Causality—via state normalisation—then gives:



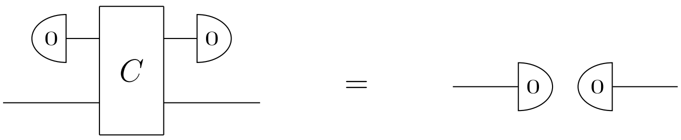
□

6.7 Proof of theorem 6.16

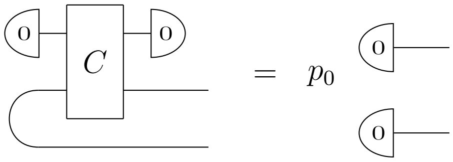
Proof. Choose the dynamically faithful state to satisfy



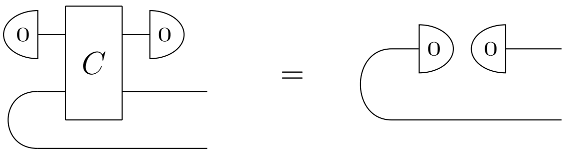
where $p_i \in [0,1]$ and $\sum_i p_i = 1$, which can always be achieved without loss of generality (see theorem 6 and corollary 9 from [15]). We first show that C satisfies



Indeed, uniqueness of measurement (both of the following states give probability p_0 for $(0|(0|$, and probability zero for each of $(0|(1|, (1|(0|$, and $(1|(1|)$ implies

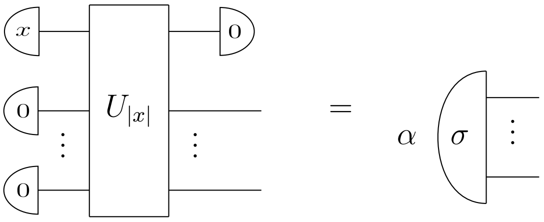


From our choice of dynamically faithful state, it then follows that

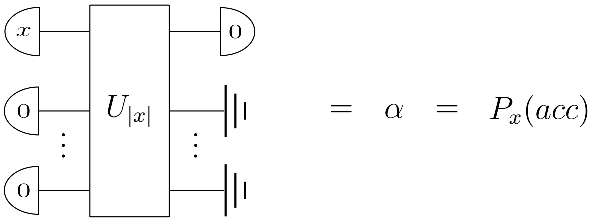


A strengthened version of dynamical faithfulness from [144] gives the required result.

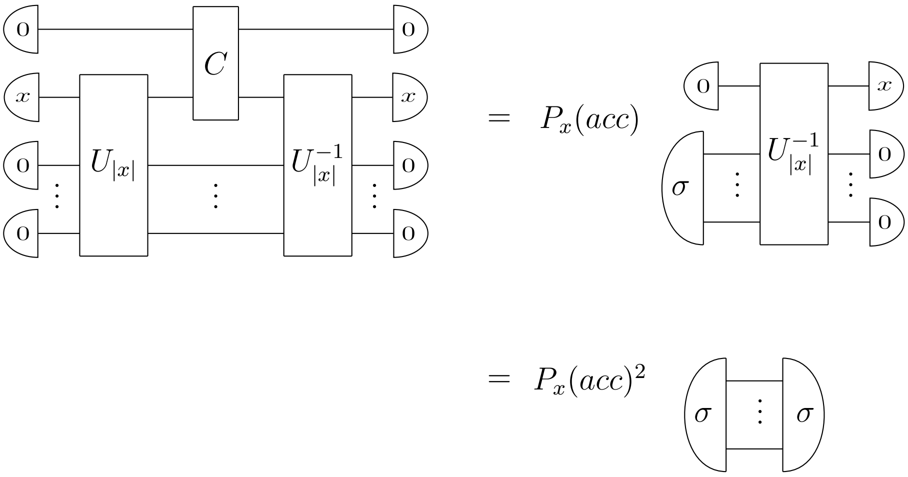
Next, note that the conjunction of purity preservation and the fact that reversible transformations map pure states to pure states gives



where $|\sigma)$ is a normalised pure state and $\alpha \in [0, 1]$. Our choice of acceptance condition, and the fact that U is a dilation of the algorithm A, results in



Combining these two results now gives



$$= \; P_x(acc)^2$$

where the last two lines follow from self-duality and strong symmetry.

Now, by amplifying the acceptance probability of the original algorithm A, we can ensure that when x is in the language we have $P_x(acc) \geq 1 - 2^{-p(|x|)}$ for an arbitrary polynomial $p(|x|)$. Hence it follows that $P_x(acc)^2 \geq 1 - 2^{-p(|x|)+1}$. Choosing $p(|x|) = q(|x|) + 1$ completes the proof. □

6.8 Conclusion

The key result of this chapter was to provide a set of physical principles that are sufficient for the existence of reversible controlled transformations. Such transformations are central to our understanding of quantum computing, information processing and thermodynamics. Moreover, these were shown to guarantee the existence of a generalised phase kick-back mechanism, which, in the quantum case, forms a fundamental component of almost all algorithms. These physical principles are defining characteristics of information: independence of encoding medium; propagation from present to future; and conservation at a fundamental level. It would therefore be surprising if these principles were not necessary primers for information processing. These results provide the tools for an exploration of the structure of computational algorithms—and how they connect to physical principles—in operational theories.

We developed a framework that connects higher-order interference and phase transformations, generalising the intimate connection between phase and interference witnessed in quantum theory. These 'higher-order' phases are accessible via our generalised kick-back mechanism. Given two assumptions which connect the operational theory to a physical description of particles, these higher-order phases were shown to give rise to exotic particle types. Additionally, using the controlled transformations to define an oracle model of computation, we conjectured that these higher-order phases may allow for the solution of problems intractable even on a quantum computer. Computational problems that may be susceptible to efficient solution by generalised phase kick-back include the n-collision problem, and the non-abelian hidden subgroup problem. Discovering that higher-order interference leads to 'unreasonable' computational power may provide a reason 'why' quantum theory is limited in its interference behaviour—in the same way that implausible communication complexity is thought to limit quantum non-locality [5]. In the next chapter, we use the tools developed in the current chapter to investigate whether higher-order interference provides a computational advantage over quantum theory in the search problem.

In section 6.2.1 it was shown that, to observe the exotic particle types corresponding to higher-order phases, there must be distinct ways to swap particles. As we live in a topologically trivial three dimensional space, there is only one topologically distinct way to swap point particles. This can either be seen as evidence of *why* quantum theory is limited to only second-order interference, or evidence that such particle types must have non-trivial structure, similar to toroidal anyons [137]—which are constructed from a solenoid ring with an attached charge—or closed strings [138].

Finally, reference [139] has shown that thermodynamic work can be extracted from quantum coherences—2nd-order phases in our language. This raises this the question of whether one can extract work more efficiently using higher-order phases? If such efficiencies are in contention with thermodynamic principles this could provide a reason 'why' quantum theory has limited interference. Initial investigations into formulating a consistent thermodynamics in operational theories have been reported in [142, 143, 144]. The framework and results presented here may therefore have implications for thermodynamics, information processing, and how each arises in a unified manner from physical principles.

Chapter 7

Higher-order interference doesn't help in searching for a needle in a haystack[*]

Grover's algorithm [149] provides the optimal quantum solution to the search problem and is one of the most versatile and influential quantum algorithms. The search problem—in its simplest form—asks one to find a single "marked" item from an unstructured list of N elements by querying an oracle which can recognise the marked item. The importance of Grover's algorithm stems from the ubiquitous nature of the search problem and its relation to **NP**-complete problems [44]. Classical computers require $O(N)$ queries to solve this problem, but quantum computers—using Grover's algorithm—only require $O(\sqrt{N})$ queries. Quantum interference between computational paths has been posited [84] as a key resource behind this computational "speed-up". However, as we discussed in chapter 5, there is a limit to this interference—at most pairs of paths can ever interact in a fundamental way [124, 125]. Could more interference imply more computational power?

Recall from the previous two chapters that Sorkin has defined a hierarchy of possible interference behaviours—currently under experimental investigation [131, 132, 153]—where classical theory is at the first level of the hierarchy and quantum theory belongs to the second. Informally, the order in the hierarchy corresponds to the number of paths that have an irreducible interaction in a multi-slit experiment. To get a greater understanding of the role of interference in computation, we consider how Grover's speed-up depends on the order of interference in a theory.

As we saw in chapter 5, restriction to the second level of this hierarchy implies many "quantum-like" features, which, at first glance, appear to be unrelated to inter-

[*] With apologies to L. Grover [149]

ference. As mentioned previously, such interference behaviour restricts correlations [127] to the "almost quantum correlations" discussed in [75], and bounds contextuality in a manner similar to quantum theory [74, 73]. This, in conjunction with interference being a key resource in the quantum speed-up, suggests that post-quantum interference may allow for a speed-up over quantum computation.

Surprisingly, we show that this is not the case—at least from the point of view of the search problem. We consider generalised probabilistic theories satisfying the principles of causality, purification, purity preservation, and strong symmetry, which, as we saw in the previous chapter, when combined with the existence of pure and perfectly distinguishable states, are sufficient for the existence of a well-defined computational oracle. Given these physical principles, we prove that a theory at level h in Sorkin's hierarchy requires $\Omega(\sqrt{N/h})$ queries to solve the search problem in the case of a single marked item. Thus, post-quantum interference does not imply a computational speed-up over quantum theory. Moreover, from the point of view of the search problem, all (finite) non-trivial orders of interference are asymptotically equivalent.

7.1 Standing assumptions

In this chapter we shall be assuming the principles of causality, purification, purity preservation, and strong symmetry, together with the existence of pure and perfectly distinguishable states. As we discussed in chapter 1, these principles do not restrict us to standard quantum theory. Indeed, real vector space quantum theory satisfies all the above principles, as does classical theory when restricted to pure states.

Recall from the previous chapter that these principles, together with the existence of (at least two) pure and perfectly distinguishable states, imply the following result, called *weak spectrality* by [72], (see [144] for a proof): for any given state $|s)$, there exists a natural number n and a set of pure and perfectly distinguishable states $\{|a^i)\}_{i=1}^n$ such that $|s) = \sum_i p_i |a_i)$ where $0 \leq p_i \leq 1$, $\forall i$ and $\sum_i p_i = 1$.

As discussed previously, this result, together with principle 5, implies the existence of a "self-dualising" [49, 72] inner product $\langle \cdot, \cdot \rangle$. That is, to every pure state $|s)$, there is associated a unique pure effect $(e^s|$, satisfying $(e^s|s) = 1$, such that: $(e^s|\cdot) = \langle |s), \cdot \rangle$. Henceforth in this chapter, we shall drop the curved brackets of the "Dirac-like" notion $|s)$ in favour of s. This is done only for stylistic reasons and makes equations involving the inner product $\langle \cdot, \cdot \rangle$ cleaner and easier to read.

The above inner product is invariant under all reversible transformations, satisfies $0 \leq \langle r, s \rangle \leq 1$ for all states r, s, $\langle s, s \rangle = 1$ for all pure states s, and $\langle s, r \rangle = 0$ if s and

r are perfectly distinguishable. It also gives rise to the norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$, satisfying $\|s\| \leq 1$ for all states s, with equality for pure states. We will make use of this norm in proving the main theorem in this chapter.

7.2 Higher-order interference in the presence of principles 2 to 5

As we saw in chapter 5, higher-order interference was initially formalised by Sorkin in the framework of Quantum Measure Theory [124] but has more recently been adapted to the setting of generalised probabilistic theories. However, given the principles of causality, purification, purity preservation, and strong symmetry, together with the existence of pure and distinguishable states, it is possible to define unique physical transformations[1] which correspond to the action of blocking certain subsets of slits. Hence, given our principles, there is a unique way to choose the effects corresponding to blocking or un-blocking slits in a multi-slit experiment. Given this structure, there is a more convenient (and equivalent, given the principles) definition of higher-order interference in terms of such transformations [72].

If there are N slits, labelled $1,\ldots,N$, these transformations are denoted P_I, where $I \subseteq \{1,\ldots,N\} := \mathbf{N}$ corresponds to the subset of slits which are not blocked. In general we expect that $P_I P_J = P_{I\cap J}$, as only those slits belonging to both I and J will not be blocked by either P_I or P_J. This intuition suggests that these transformations should correspond to projectors (i.e. idempotent transformations $P_I P_I = P_I$). Given principles 2 to 5, together with the existence of pure and perfectly distinguishable states, it was shown in [72] that this is indeed the case. Given this structure, one can define the maximal order of interference as follows [72].

Definition 7.1. *A theory satisfying causality, purification, purity preservation, and strong symmetry, together with the existence of a set of N pure and perfectly distinguishable states for every natural number N, has maximal order of interference h if, for any $N \geq h$, one has:*

$$\mathbb{1}_N = \sum_{\substack{I \subseteq \mathbf{N} \\ |I| \leq h}} \mathcal{C}\,(h, |I|, N)\, P_I$$

[1] Barnum *et al.* prove in [72] that the conjunction of strong symmetry and weak spectrality implies such physical transformations exist.

where $\mathbb{1}_N$ is the identity on a system with N pure and perfectly distinguishable states and

$$\mathcal{C}\left(h,|I|,N\right) := (-1)^{h-|I|}\left(\begin{array}{c} N-|I|-1 \\ h-|I| \end{array}\right)$$

The factor $\mathcal{C}\left(h,|I|,N\right)$ in the above definition (which is proportional to a binomial coefficient) corrects for the overlaps that occur when different combinations of slits are blocked. Note that, for the case $h=N$, definition 7.1 reduces to the expected expression of $\mathbb{1}_h = P_{\{1,\ldots,h\}}$ i.e. the identity is given by the projector with all slits open. The case of $N=h+1$ corresponds to $\mathcal{C}\left(h,|I|,h+1\right)=(-1)^{h-|I|}$, which is the situation considered in chapters 5 and 6, as well as the one most commonly discussed in the literature [124, 78]. In the specific case of quantum theory, definition 7.1 reduces to

$$\mathbb{1}_N = \sum_{i<j} P_{\{ij\}} - (N-2)\sum_i P_{\{i\}},$$

where $P_{\{ij\}}$ sends all but the ii, jj, ij, and ji entries of a given $N\times N$ density matrix to zero, and $P_{\{i\}}$ sends all but the ii entry of a given $N\times N$ density matrix to zero.

Rather than work directly with these physical projectors, it is mathematically more convenient to work with the (generally) unphysical maps corresponding to projectors onto the "coherences" of a state. For example, in the case of a qutrit, the projector $P_{\{0,1\}}$ projects onto a two dimensional subspace:

$$P_{\{0,1\}} :: \begin{pmatrix} \rho_{00} & \rho_{01} & \rho_{02} \\ \rho_{10} & \rho_{11} & \rho_{12} \\ \rho_{20} & \rho_{21} & \rho_{22} \end{pmatrix} \mapsto \begin{pmatrix} \rho_{00} & \rho_{01} & 0 \\ \rho_{10} & \rho_{11} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

whilst the coherence-projector $\omega_{\{0,1\}}$ projects only onto the coherences in that two dimensional subspace:

$$\omega_{\{0,1\}} :: \begin{pmatrix} \rho_{00} & \rho_{01} & \rho_{02} \\ \rho_{10} & \rho_{11} & \rho_{12} \\ \rho_{20} & \rho_{21} & \rho_{22} \end{pmatrix} \mapsto \begin{pmatrix} 0 & \rho_{01} & 0 \\ \rho_{10} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

That is, $\omega_{\{0,1\}}$ corresponds to the linear combination of projectors: $P_{\{0,1\}} - P_{\{0\}} - P_{\{1\}}$.

There is a coherence-projector ω_I for each subset of slits $I \subseteq \mathbf{N}$, defined in terms of the physical projectors:

$$\omega_I := \sum_{\widetilde{I}\subseteq I} (-1)^{|I|+|\widetilde{I}|} P_{\widetilde{I}}.$$

These have the following useful properties:

Lemma 7.2. *An equivalent definition of the maximal order of interference, h, is:*

$$\mathbb{1}_N = \sum_{I,|I|=1}^{h} \omega_I, \text{ for all } N \geq h.$$

Proof. In a theory with maximal order of interference h one has

$$\mathbb{1}_N = \sum_{\substack{I \subseteq \mathbf{N} \\ |I| \leq h}} \mathcal{C}\left(h, |I|, N\right) P_I.$$

Thus, showing $\mathbb{1}_N = \sum_{|I|=1}^{h} \omega_I$ reduces to showing

$$\sum_{|I|=1}^{h} \omega_I = \sum_{\substack{I \subseteq \mathbf{N} \\ |I| \leq h}} \mathcal{C}\left(h, |I|, N\right) P_I.$$

As $\omega_I := \sum_{\widetilde{I} \subseteq I} (-1)^{|I|+|\widetilde{I}|} P_{\widetilde{I}}$, we just have to count the number of P_I's that appear as we sum over $|I|$. For some fixed I, this is just

$$\sum_{\alpha=|I|}^{h} (-1)^{\alpha-|I|} \left(\begin{array}{c} N-|I| \\ \alpha-|I| \end{array} \right).$$

By expanding and rearranging this, one can straightforwardly (if tediously) show that this equals $\mathcal{C}\left(h, |I|, N\right)$, and we are done. □

The above lemma implies that any state (or indeed, any vector in the vector space generated by the states) can be decomposed as $s = \sum_{I,|I|=1}^{h} s_I$, where $s_I := \omega_I s$. That is, states belonging to theories which satisfy our principles and lie at level h in Sorkin's hierarchy decompose in a form reminiscent of rank-h tensors. Hence, theories satisfying our principles bear some resemblance to the theory of Density Cubes from chapter 5.

Lemma 7.3. *"Coherences are orthogonal":*

(i) $\omega_I \omega_J = \delta_{IJ} \omega_I$ *for all* I, J, *and*

(ii) $\|s\|^2 = \sum_I \|\omega_I s\|^2$.

Proof of part (i). From the definition of ω_I, it follows that

$$\omega_I \omega_J = (-1)^{|I|+|J|} \sum_{\widetilde{I} \subseteq I} \sum_{\widetilde{J} \subseteq J} (-1)^{|\widetilde{I}|+|\widetilde{J}|} P_{\widetilde{I}} P_{\widetilde{J}}$$

$$= (-1)^{|I|+|J|} \sum_{\widetilde{K} \subseteq I \cap J} \mathcal{D}\left(I, J, \widetilde{K}\right) P_{\widetilde{K}}$$

where $\mathcal{D}\left(I, J, \widetilde{K}\right)$ is the number of distinct pairings of $\widetilde{I}$ and $\widetilde{J}$ such that $\widetilde{I} \cap \widetilde{J} = \widetilde{K}$ and $|\widetilde{I}| + |\widetilde{J}|$ is even, minus the number of distinct pairings where $\widetilde{I} \cap \widetilde{J} = \widetilde{K}$ and $|\widetilde{I}| + |\widetilde{J}|$ is odd. It will now be shown that

$$\mathcal{D}\left(I, J, \widetilde{K}\right) = \begin{cases} 0 & \text{if } I \neq J \\ (-1)^{|I|+|\widetilde{K}|} & \text{if } I = J \end{cases}$$

For the $I \neq J$ case fix some particular $i \in I$ such that $i \notin J$ and consider some $\widetilde{I} \subseteq I, \widetilde{J} \subseteq J$ such that $\widetilde{I} \cap \widetilde{J} = \widetilde{K}$. If $x \notin \widetilde{I}$ alter $\widetilde{I}$ by adding i, otherwise alter $\widetilde{I}$ by removing x. This procedure turns each even $|\widetilde{I}| + |\widetilde{J}|$, odd. We have thus shown that for each $\widetilde{I} \subseteq I$ and $\widetilde{J} \subseteq J$ such that $\widetilde{I} \cap \widetilde{J} = \widetilde{K}$ and $|\widetilde{I}| + |\widetilde{J}|$ is even, there exists an $\widetilde{I}' \subseteq I$ such that $\widetilde{I}' \cap \widetilde{J} = \widetilde{K}$ and $|\widetilde{I}'| + |\widetilde{J}|$ is odd, and vice versa. Thus the number of distinct pairings of $\widetilde{I}$ and $\widetilde{J}$ such that $\widetilde{I} \cap \widetilde{J} = \widetilde{K}$ and $|\widetilde{I}| + |\widetilde{J}|$ is even is equal to the number of distinct pairings of $\widetilde{I}$ and $\widetilde{J}$ such that $\widetilde{I} \cap \widetilde{J} = \widetilde{K}$ and $|\widetilde{I}| + |\widetilde{J}|$ is odd, and so $\mathcal{D}\left(I, J, \widetilde{K}\right) = 0$ when $I \neq J$.

For the $I = J$ case we can make a similar argument by picking some $i \in I, i \notin \widetilde{J}$ except for when $\widetilde{J} = J = I$. This case gives an excess ±1 depending on whether $|J| + |\widetilde{K}|$ is odd or even, implying $\mathcal{D}\left(I, J, \widetilde{K}\right) = (-1)^{|I|+|\widetilde{K}|}$ when $I = J$.

This immediately gives $\omega_I \omega_J = 0$ if $I \neq J$ and,

$$\omega_I \omega_I = (-1)^{2|I|} \sum_{\widetilde{K} \subseteq I} (-1)^{|I|+|\widetilde{K}|} P_{\widetilde{K}} = \omega_I$$

if $I = J$. □

Proof of part (ii). To prove the lemma, we need the fact that the ω_I's are self-dual $\omega_I^\dagger = \omega_I$, where the † is defined by the the self-dualising inner-product as: $\langle \cdot, \omega_I \cdot \rangle = \langle \omega_I^\dagger \cdot, \cdot \rangle$. Recalling that the ω_I's correspond to linear combinations of the P_I's, this follows immediately from self-duality of the projectors P_I, which is shown to follow from weak spectrality and strong symmetry in [72]. We now have

$$\|s\|^2 = \langle s, s \rangle = \langle \sum_I \omega_I s, \sum_J \omega_J s \rangle$$

$$= \sum_{I,J} \langle \omega_I s, \omega_J s \rangle = \sum_{I,J} \langle s, \omega_I^\dagger \omega_J s \rangle$$

$$= \sum_{I,J} \langle s, \omega_I \omega_J s \rangle = \sum_{I,J} \delta_{IJ} \langle s, \omega_I s \rangle$$

where the last equality follows from the orthogonality of the ω_I's. Finally

$$\|s\|^2 = \sum_I \langle s, \omega_I s\rangle = \sum_I \langle s, \omega_I^2 s\rangle = \sum_I \langle \omega_I s, \omega_I s\rangle = \sum_I \|\omega_I s\|^2$$

□

7.3 Setting up the problem

In the standard search problem, one is asked to find a unique "marked" item from among a large collection of items in some unstructured list. The items are indexed $1, \ldots, N$ and one has access to an oracle, which, when asked whether item i is the marked item, denoted x, returns the answer "yes" or "no". Informally, the search problem asks for the minimal number of queries to this oracle required to find x in the worst case.

In the standard bra-ket formalism of quantum theory, this oracle corresponds to a controlled unitary transformation U, defined by its action on the (product) computational basis: $U|i\rangle|q\rangle = |i\rangle|q \oplus f(i)\rangle$, where $|i\rangle$ is the index, or control, register, $|q\rangle$ is the target register, $\oplus$ denotes addition modulo 2 and $f : \{1, \ldots, N\} \to \{0, 1\}$ satisfies $f(i) = 1$ if and only if $i = x$. Inputting $|-\rangle$ into the target register results in a phase being "kicked-back" to the control register: $U|i\rangle|-\rangle = (-1)^{f(i)}|i\rangle|-\rangle$. Discarding the target register reduces the action of the oracle to applying the phase transformation $O_x|i\rangle = (-1)^{f(i)}|i\rangle$. Changing to the density matrix formalism, we see that this phase oracle, whose action on states ρ is now denoted by $\mathcal{O}_x\rho$, acts as the identity on all entries of a given density matrix except for the off diagonal elements $\{\rho_{xi}, \rho_{ix}\}_i$, to which it adds a '$-$'.

Theorem 6.12 from chapter 6 showed that the conjunction of principles 2 to 5 with the existence of pure and perfectly distinguishable states implies the existence of reversible controlled transformations. As we saw in section 6.2.2 from chapter 6, such transformations can be used to define oracles in a manner analogous to quantum theory. Moreover, theorem 6.15 showed every controlled transformation gives rise to a "kicked-back" reversible phase transformation on the control system [89]. Thus—as in quantum theory—from the point of view of querying the oracle, we can reduce all considerations involving the controlled transformation to those involving the kicked-back phase.

To highlight the role of interference in searching an unstructured list, we describe the action of querying the oracle in terms of the physically motivated set-up of N-slit experiments. Consider first the quantum case. Note that an N-slit experiment

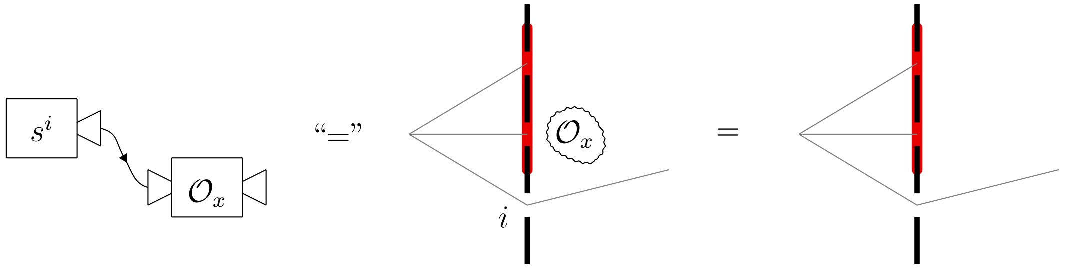


Fig. 7.1: Querying an oracle using a multi-slit experiment

defines a set of N pure and perfectly distinguishable states $|i\rangle\langle i|$, each of which can be associated to a distinct element in the N item list. Querying the oracle about item i is equivalent to applying the oracle transformation to state $|i\rangle\langle i|$. In quantum theory, preparing such a state can be achieved by passing a uniform superposition through the N-slit experiment with all but the ith slit blocked. The oracle can be implemented by placing a phase shifter behind slit x. Querying the oracle in a superposition of states can then be achieved by varying which slits are blocked. This is illustrated in Fig. 7.1. As discussed previously, the physical act of blocking slits is represented by the projectors P_I. The action of the quantum oracle can thus be rephrased in terms of these projectors: i) $\mathcal{O}_x P_I = P_I$, if $x \notin I$ or $|I| = 1$ and, ii) $\mathcal{O}_x$ can act non-trivially on projectors P_I with $x \in I$ and $|I| > 1$, but must satisfy $\mathcal{O}_x P_I = P_I \mathcal{O}_x$, for all P_I, which corresponds to the fact that a quantum oracle does not "create" or "destroy" coherence between states passing through different slits.

By direct analogy with the quantum case, we define the oracle which encodes the search problem in theories satisfying our principles as follows. Note that in this thesis we only deal with the case of a single marked item.

Definition 7.4. *A reversible transformation is a* search oracle*, denoted $\mathcal{O}_x$, if and only if:*

i) $\mathcal{O}_x P_I = P_I$ *for all* $x \notin I$ *or* $|I| = 1$ *and,*

ii) $\mathcal{O}_x P_I = P_I \mathcal{O}_x$*, for all* P_I*.*

In definition 7.4, the requirement $\mathcal{O}_x P_I = P_I \mathcal{O}_x$, for all P_I, is quite natural. This requirement ensures that one cannot gain any information about item i when querying the oracle using a state with no support on i, i.e. a state s such that $P_I s = s$ where $i \notin I$. Moreover, the above constraint does not restrict the power of a search oracle. Indeed, in the search problem, one wants to determine the minimal number of queries to $\mathcal{O}_x$ required to find the marked item in the worst case. In one query, the most the oracle can "move" a state—relative to the norm $\|\cdot\|$—is to map it to a state which is perfectly distinguishable from the original. This follows from the fact that, for any two states ρ, σ, we have

$$\|\rho - \sigma\|^2 = \langle \rho - \sigma, \rho - \sigma \rangle \leq 2(1 - \langle \rho, \sigma \rangle) \leq 2,$$

with equality when ρ and σ are pure and perfectly distinguishable. Let F_I be the set of states left invariant by the projector P_I (note that these F_I are in fact *convex faces*, see [72]). The condition $P_I \mathcal{O}_x = \mathcal{O}_x P_I$ implies $\mathcal{O}_x(F_I) \subseteq F_I$, i.e. the oracle leaves the set F_I invariant. Now, one can show that each F_I defines a state space which satisfies[2] our principles (again, see [72]). Hence, for any given state in F_I with $|I| > 1$, there exists another state in the same set which is perfectly distinguishable from it. Thus, enforcing the property $P_I \mathcal{O}_x = \mathcal{O}_x P_I$ does not restrict the power of the oracle.

In an arbitrary theory, it may not be the case that a transformation satisfying definition 7.4 and acting non-trivially on P_I, with $x \in I$, exists. This is not an issue as in such theories we cannot even define the search problem, let alone show it can be solved using fewer queries than quantum theory. Henceforth, we shall assume the existence of a search oracle in any theory we consider.

Given the definition of coherence-projectors ω_I we can equivalently write definition 7.4 as: $\mathcal{O}_x \omega_I = \omega_I$, for $x \notin I$ or $|I| = 1$, and $\mathcal{O}_x \omega_I = \omega_I \mathcal{O}_x$, for all I. Indeed, in the quantum case, the action of the oracle can be equivalently described as:

$$\mathcal{O}_x \omega_I = \begin{cases} \omega_I & \text{if } x \notin I \text{ or } |I| = 1 \\ -\omega_I & \text{otherwise.} \end{cases}$$

We can now formally state the search problem for a single marked item—defined for the quantum case in [1, 154, 148]—as:

Search Problem. *Given an N element list with search oracle $\mathcal{O}_x$ and an arbitrary collection of reversible transformations $\{G_i\}$, what is the minimal $k \in \mathbb{N}$ such that $G_k \mathcal{O}_x G_{k-1} \dots G_1 \mathcal{O}_x s$ can be found, with probability greater than $1/2$, to be in the state x, for arbitrary input state s, averaged over all possible marked items?*

7.4 Main result

Theorem 7.5. *In theories satisfying causality, purification, purity preservation, and strong symmetry, with finite maximal order of interference h, and in which there*

[2] Actually, reference [72] proved that in any theory satisfying weak spectrality and strong symmetry the convex faces define a sub-theory which also satisfies weak specrality and strong symmetry, which is all that is needed here.

exists a set of n pure and perfectly distinguishable states for all positive integers n, the number of queries needed to solve the search problem is $\Omega(\sqrt{N/h})$.

Proof of theorem 7.5. The basic idea is based on the proof of the quantum case presented in [1, 148, 154]. Let

$$\begin{aligned} s_k^x &= G_k \mathcal{O}_x G_{k-1} \dots G_1 \mathcal{O}_x s, \\ s_k &= G_k G_{k-1} \dots G_1 s, \end{aligned}$$

where G_i is some reversible transformation from the theory, and define

$$D_k = \sum_x \|s_k^x - s_k\|^2.$$

It will be shown that, for $\langle x, s_k^x \rangle \geq 1/2$, we have $cN \leq D_k \leq 4hk^2$, where c is any constant less than $\left(\sqrt{2}-1\right)^2$, from which the result $k \geq O\left(\sqrt{\frac{N}{h}}\right)$ follows. The upper bound will now be shown using induction, deferring a derivation of the lower bound left to the end of the current proof.

We have

$$\begin{aligned} D_{k+1} &= \sum_x \|G_{k+1}\left(\mathcal{O}_x s_k^x - s_k\right)\|^2 = \sum_x \|\mathcal{O}_x s_k^x - s_k\|^2 \\ &= \sum_x \|\mathcal{O}_x\left(s_k^x - s_k\right) + \left(\mathcal{O}_x - \mathbb{1}\right) s_k\|^2 \\ &\leq \sum_x \|s_k^x - s_k\|^2 + 2\sum_x \|\mathcal{O}_x\left(s_k^x - s_k\right)\| \|\left(\mathcal{O}_x - \mathbb{1}\right) s_k\| + \sum_x \|\left(\mathcal{O}_x - \mathbb{1}\right) s_k\|^2 \\ &\leq D_k + 2\sqrt{D_k \sum_x \|\left(\mathcal{O}_x - \mathbb{1}\right) s_k\|^2 + \|\left(\mathcal{O}_x - \mathbb{1}\right) s_k\|^2} \\ &\leq \left(\sqrt{D_k} + \sqrt{\sum_x \|(\mathbb{1} - \mathcal{O}_x) s_k\|^2}\right)^2, \end{aligned}$$

which follows from the triangle inequality, the Cauchy-Schwarz inequality, and the fact the norm is invariant under reversible transformations.

The quantity $\sum_x \|(\mathbb{1} - \mathcal{O}_x) s_k\|^2$—which can be thought of as how much some state is “moved” in a single query, averaged over all possible marked items x—is the only theory dependent quantity that features in this proof. We upper bound it as follows:

$$\begin{aligned}
&\sum_x \|(\mathbb{1} - \mathcal{O}_x)s_k\|^2 \\
&= \sum_x \sum_I \|(\mathbb{1} - \mathcal{O}_x)\omega_I s_k\|^2 = \sum_x \sum_{\substack{I \\ |I| > 1 \\ x \in I}} \|\omega_I(\mathbb{1} - \mathcal{O}_x)s_k\|^2 \\
&\leq \sum_x \sum_{\substack{I \\ |I| > 1 \\ x \in I}} (\|\mathbb{1}\omega_I s_k\| + \|\mathcal{O}_x\omega_I s_k\|)^2 \leq \sum_x \sum_{\substack{I \\ |I| > 1 \\ x \in I}} 4\|\omega_I s_k\|^2,
\end{aligned}$$

where the first line follows from lemma 7.2, lemma 7.3, and the definition of the search oracle $\mathcal{O}_x$, and second from the triangle inequality and the fact that the norm is invariant under reversible transformations. We need to know how many times each $\|\omega_I s_k\|^2$ appears when we sum over the marked item x. Each given $I = \{i_1, i_2, \dots, i_{|I|}\}$ will appear $|I|$ times as we sum over x, one for every time i_j is the marked item. Thus

$$\begin{aligned}
&\sum_x \|(\mathbb{1} - \mathcal{O}_x)s_k\|^2 \leq \sum_{\substack{I \\ |I| > 1}} 4|I|\|\omega_I s_k\|^2 \\
&\leq 4\sum_I |I|\|\omega_I s_k\|^2 \leq 4h\sum_I \|\omega_I s_k\|^2 = 4h\|s_k\|^2 \leq 4h.
\end{aligned}$$

The second line follows from $\sum_{|I|=1} \|\omega_I s_k\|^2 \geq 0$, lemma 7.3, $\|s_k\| \leq 1$, and $|I| \leq h$, for all I. We thus have: $D_{k+1} \leq \left(\sqrt{D_k} + \sqrt{4h}\right)^2$. Assuming that $D_k \leq 4hk^2$ gives us $D_{k+1} \leq 4h(k+1)^2$, from which the result follows via induction.

We assume that $\langle x, s_k^x\rangle \geq 1/2$ for all x, so a measurement of s_k^x yields a solution to the search problem with probability at least $1/2$. Let $E_k = \sum_x \|s_k^x - x\|^2$ and $F_k = \sum_x \|s_k - x\|^2$. It follows that

$$\text{i) } E_k = \sum_x 2(1 - \langle x, s_k^x\rangle) \leq \sum_x 2(1 - 1/2) \leq N \text{ and,}$$

$$\text{ii) } F_k \geq 2\left(N - \|s_k\|\sqrt{\left\langle \sum_x x, \sum_y y \right\rangle}\right) \geq 2\left(N - \sqrt{N}\right)$$

where ii) follows from the Cauchy-Schwarz inequality, $\|s_k\| \leq 1$ and $\langle x, y\rangle = \delta_{xy}$. Using the reverse triangle inequality and the Cauchy-Schwarz inequality, it follows that $D_k \geq \left(\sqrt{F_k} - \sqrt{E_k}\right)^2$. Combining this with the upper bound on E_k and the lower bound on F_k, we have that $D_k \geq cN$, for sufficiently large N, where c is any constant less than $\left(\sqrt{2} - 1\right)^2 \approx 0.17$.

We thus have $cN \leq D_k \leq 4hk^2$, from which the desired result follows. □

7.5 Discussion

In this chapter, we considered theories satisfying certain natural physical principles which are sufficient for the existence of controlled transformations and a phase kick-back mechanism, necessary features for a well-defined search oracle. Given these physical principles, we proved that a theory with maximal order of interference h requires $\Omega(\sqrt{N/h})$ queries to this oracle to find a single marked item from some N-element list. This result is somewhat surprising given that one might expect more interference to imply more computational power. Our main result can be thought of as a derivation of Grover's quadratic lower bound from simple physical principles—the computational analogue of deriving Cirel'son's bound on quantum correlations from physical principles such as Information Causality [77] or Local Orthogonality [76].

Further work will focus on determining sufficient physical principles for there to exist an algorithm that achieves the quadratic lower bound derived here. Moreover, as theories satisfying our five physical principles appear 'quantum-like'—at least from the point of view of the search problem—investigating interference behaviour in them may inform current experiments searching for post-quantum interference.

Finally, recent work has also investigated Grover's algorithm from the point of view of post-quantum theories [146, 147]. These works considered modifications of quantum theory which allow for superluminal signalling and cloning of states. In contrast, the generalised probabilistic theory framework employed here allowed us to investigate Grover's lower bound in alternate theories that are physically reasonable and which, for example, do not allow for superluminal signalling or cloning. Based on this, researchers interested in exploring post-quantum theories—such as those arising from the black hole firewall and information loss paradoxes—may find this framework an appealing arena in which to explore modifications of quantum theory. Indeed, preliminary work has already begun in this direction [114].

Summary and future work

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

Excerpt from "Little Gidding"
T. S. Eliot

This thesis has explored some connections between computation and physical principles in the framework of generalised probabilistic theories. The main focus has been on understanding how simple physical principles bound the power of different computational paradigms. In chapter 2 we showed that in any theory satisfying tomographic locality the class of problems that can be solved efficiently is contained in the (slightly obscure) classical complexity class **AWPP**—the best known bound on the power of quantum computation [45]. Tomographic locality was needed in order to ensure that solutions to computationally difficult problems couldn't be encoded in "global degrees of freedom" which—in the absence of tomographic locality—could arise when constructing computational circuits in certain theories. In chapter 3 we showed this containment to be tight by constructing a concrete theory (albeit in a modified framework) whose computational power *exactly* equals **AWPP**.

The above result raises the question of whether quantum theory is powerful for computation in the space of all theories. Given a specific theory, under what conditions can computation in this theory be simulated by a quantum computer? Put differently, can the "quantum region" of the computational landscape illustrated in Fig. 3.3 be characterised in terms of physical principles alone? Such a characterisation would deepen our understanding of quantum computation and its ultimate limitations. Moreover, it would provide a theory-independent characterisation of the class of problems a quantum computer can solve efficiently.

One approach to such a characterisation would be to investigate which physical principles—when satisfied by a theory **G**—ensure a given **BQP**-complete problem is **BGP**-hard. An example of such a problem is matrix inversion: the ability to invert sparse, well-conditioned matrices, together with the ability to simulate a single system measurement, is a **BQP**-complete problem. This was first shown by Harrow, Hassidim, and Lloyd in their paper [155]. Their work provided an efficient quantum algorithm for solving (certain types of) linear systems of equations [155]. Determining which physical principles ensure the process of calculating the outcomes of efficient computations can be reduced to solving the matrix inversion problem would provide a class of theories which can be simulated by quantum theory.

In chapter 4 we investigated how simple physical principles bound the power of two different computational paradigms which combine computation and communication in a non-trivial fashion: computation with advice and interactive proof systems. We showed that in theories satisfying the principle of strong symmetry the power of computation with advice is contained in the class **PP/poly**. Moreover, we showed that in Boxworld—which does not satisfy strong symmetry—the power of computation with advice is unbounded. Additionally, we proved that the power of simple interactive proof systems in theories satisfying tomographic locality is contained in $\mathbf{A_0PP}$ (which is itself contained in **PP**). Finally, we argued that these results could, in some sense, be seen to illuminate the (still unresolved!) "trade-off conjecture". As such, further explorations of computational paradigms which combine both computation and communication in non-trivial ways could help us understand quantum theory's place in the broad framework of generalised probabilistic theories.

Over the course of chapters 5, 6, and 7, we explored the structure of computational algorithms in the generalised probabilistic theory framework. Moreover, we investigated whether certain algorithmic advantages are directly related to simple physical principles, focusing on whether the existence of post-quantum interference implies a speed-up over quantum computation. In chapter 6 we showed that, in any theory satisfying causality, purification, purity preservation, and strong symmetry, and in which there are sufficient distinguishable states, reversible controlled transformations exist. Moreover, each reversible controlled transformation has a phase kick-back mechanism. We then used these results to show that non-trivial interference behaviour is a general resource for post-classical computation.

In chapter 7 we considered how Grover's speed-up depends on the order of interference in a theory. We restricted to theories satisfying causality, purification, purity preservation, and strong symmetry, which, as mentioned above, are sufficient for the existence of reversible controlled transformations and hence a well-defined search oracle. In such theories we showed that the quadratic lower bound to the search problem holds regardless of the order of interference[3]. That is, we proved that a theory at level h in Sorkin's hierarchy requires $\Omega(\sqrt{N/h})$ queries to solve the search problem. Hence, at least from the point of view of the search problem, post-quantum interference is not a resource for post-quantum computation.

The derivation of the quadratic lower bound to the search problem from causality, purification, purity preservation, strong symmetry, the existence of sufficient distinguishable states, and a finite order of interference raises the question of whether these physical principles are sufficient for the existence of an efficient algorithm which achieves this lower bound[4]. A quantum search algorithm based on Hamiltonian simulation, presented in chapter 6 of the textbook by Nielsen and Chuang [1], may be more directly generalisable to theories satisfying the above principles than Grover's original construction. Indeed, in [1] they consider a Hamiltonian H consisting of projectors onto the marked item $|x\rangle$ and the initial input state $|\psi\rangle = \alpha|x\rangle + \beta|y\rangle$, with $|y\rangle$ orthogonal to $|x\rangle$ and $\alpha^2 + \beta^2 = 1$, respectively. That is, they consider the Hamiltonian $H = |x\rangle\langle x| + |\psi\rangle\langle\psi|$. Evolving the initial input state under this Hamiltonian for time t results in

$$\exp(-itH)|\psi\rangle = \cos(\alpha t)|\psi\rangle - i\sin(\alpha t)|x\rangle.$$

Hence, measuring the system in the $\{|x\rangle, |y\rangle\}$ basis at time $t = \pi/2\alpha$ yields outcome $|x\rangle$ with probability one. If the initial state was a uniform superposition over the orthonormal basis containing $|x\rangle$, then the required evolution time is $t = \pi\sqrt{N}/2$, where N is the size of the system (or equivalently, the number of elements in the list being searched).

One might wonder why there is no mention of a search oracle in the above discussion. The oracle comes into play when constructing a quantum circuit to simulate the above Hamiltonian evolution. As the above Hamiltonian depends on the marked item, the quantum circuit simulating it must query the search oracle a number of times proportional to the evolution time [1]. In this specific case, an efficient Hamiltonian simulation requires $O(\sqrt{N})$ queries to the oracle, yielding an optimal quantum algorithm (up to constant factors) for the search problem. Recently, Barnum et al. [72] have introduced a physical principle, termed "energy observability", which implies the existence of a continuous time evolution and ensures that the generator of such

[3] As long as the order of interference is finite.

[4] Recall that classical theory violates purification and so is ruled out by these principles.

an evolution—a generalised "Hamiltonian"—is associated to an appropriate observable, which is a conserved quantity—the generalised "energy" of the evolving system. Recall from chapter 7 that the principles we have discussed—causality, purification, purity preservation, and strong symmetry—were sufficient to ensure that projectors onto arbitrary states correspond to allowed transformations. Hence, our previous principles, together with Barnum et al.'s energy observability, may be sufficient to run the above quantum search algorithm, hence providing a theory independent description of an optimal (up to constant factors) search algorithm.

While the results in this thesis may have deepened our understanding of quantum computers and their limits, they have not explicitly resulted in any practical applications in the same way that studying Boxworld type correlations led to the development of device-independent cryptography [10]. Can studying computation in general theories different from quantum theory result in practical applications? One potential avenue for this is blind and verified delegated computation [156].

Consider the situation where a computationally bounded client wants to delegate her computation to a server with access to a full quantum computer. The protocol for blind computation provided in [156] ensures that the client can have a server carry out a quantum computation for her such that the client's inputs, outputs, and computation remain perfectly private. Hence, a malicious server cannot learn any of the client's information, and all the client needs to be able to do is prepare single qubit states and send them to the server. Moreover, the security of this protocol has been shown to follow from the no-signalling principle [157]. However, while the server cannot learn the clients computation, he can still tamper with it by deviating from the client's instructions[5].

Reference [156] also provides a protocol which verifies whether the server deviated from the specified instructions while still maintaining blindness. The basic idea of the verification protocol presented in [156] is to hide "trap" qubits in the computation. The server does not know the position of the traps, and if he touches one he changes its state with a certain probability. The client checks the traps and accepts the result of the computation only if no trap is sprung. Hence to ensure any deviations are detected with high probability, the client must maximise the probability of the server touching a trap qubit if he deviates from the correct procedure. This is achieved using quantum error correcting codes: if the computation is encoded with a quantum error

[5] Indeed, the server could simply refuse to perform the client's computation. Such a refusal is immediately obvious to the client. As nothing can be done to force the server to perform the computation, we only consider situations in which the server does carry out a computation. It is then up to the client to verify if the performed computation is the one specified.

correcting code, the server has to apply global operations to alter the logical qubits, and this requirement drastically increases the probability of the server touching a trap hidden among the physical qubits. Using this scheme, the client can detect any deviations made by the server with probability exponentially close to one. However, the correctness of this verification protocol rests on the assumption that the server can only deviate from the specified instructions by using quantum dynamics. The client may not be able to detect deviations which use *post-quantum* dynamics.

This raises the question of whether the correctness of this verification protocol can be established directly from physical principles. This question is of practical interest; if quantum mechanics were ever to fail in some regime, a technologically and scientifically advanced server could potentially use post-quantum dynamics to deviate from the specified computation without necessarily springing a trap. For example, if an experiment were found to support the existence of higher-order interference, the server could deviate from the desired computation by applying one of the higher-order phase transformations introduced in chapter 6. As such transformations are undetectable on certain subsystems it is conceivable that they could alter the final outcome of the computation, yet not spring any trap along the way. Hence, as was the case for quantum key distribution before the work of Barrett, Hardy, and Kent [10], delegated computation is open to a post-quantum attack.

Some of the results in this thesis lay the groundwork for establishing the correctness of the above verification protocol from physical principles alone. Indeed, the protocol presented in [156] uses techniques from measurement-based quantum computation and quantum error correction to establish verifiability (under the assumption that quantum theory is correct) of the delegated computation. The basic idea of measurement-based computation is to apply reversible controlled transformations, phase transformations, and single qubit measurements to a large entangled state in order to perform some desired computation. In chapter 6 we showed that reversible controlled transformations exist in any theory satisfying causality, purification, purity preservation, and strong symmetry. This, in conjunction with the fact that causality and purification imply both the existence of entangled states and the ability to perform rudimentary error correction [15], suggests that a verification protocol may be derivable from these physical principles alone.

Another way to discuss verifying delegated computation is to use the language of quantum prover interactive proofs [158]. Quantum prover interactive proofs consider the case where a verifier is limited to **BPP** computations, coupled with limited quantum information processing, and has to verify proofs from a prover with the full power of **BQP**. Another way to phrase the question raised above then presents itself: can a **BPP** computer—coupled with limited quantum hardware—verify proofs from a **BGP** computer, where **G** satisfies some natural physical principles? In chapter 4 we discussed simple examples of quantum interactive proofs systems, and it would be interesting if such considerations could be brought to bear on this practically motivated problem. Indeed, if one could show that a **BPP** computer, together with limited quantum information processing, could verify proofs from an **AWPP** computer, then one would have established the ability to verify delegated computation directly from the principle of tomographic locality.

To conclude, this thesis has studied the bounds on computation imposed by simple physical principles. Moreover, we have suggested potential practical implications of such investigations. However, the main motivation for undertaking this study was foundational in nature. Understanding in a rigorous manner how the structure and limitations of computation are connected to physical principles deepens our knowledge of the structure of quantum theory—and of quantum computation in particular. Such an understanding takes us a step closer to knowing what it means to say we live in a quantum world.

Appendix A

A compendium of complexity classes

In this appendix, the definitions of complexity classes mentioned in this thesis—but not explicitly defined thus far—shall be presented. Relations between these classes are dicpicted in Fig. A.1. It is assumed that the reader has some familiarity with the definition of a Turing Machine, both the deterministic and probabilistic varieties. For a review of these definitions, see [159]. For the class **AWPP**, see definition 2.3 from chapter 2. For the class **BQP** replace **G** with **Q** in definition 2.1, also from chapter 2.

Definition A.1 (**P**)**.** *A language $\mathcal{L}$ is in **P** if and only if there exists a polynomial-time deterministic Turing machine M, such that*

1. *For every $x \in \mathcal{L}$, M accepts*

2. *For every $x \notin \mathcal{L}$, M rejects*

One can also define **P** in terms of a uniform family of poly-size Boolean circuits, see [159] for a discussion of this point.

Definition A.2 (**NP**)**.** *A language $\mathcal{L}$ is in **NP** if and only if there exist polynomials p and q, and a deterministic Turing machine M with two inputs x, y, such that*

1. *For all strings x and y, the Turing Machine M runs in time $p(|x|)$ on input of x, y*

2. *For every $x \in \mathcal{L}$, there exists a string y of length $q(|x|)$ such that, on input of x, y, M accepts*

3. *For all $x \notin \mathcal{L}$, and all strings y of length $q(|x|)$, M rejects on input of x, y*

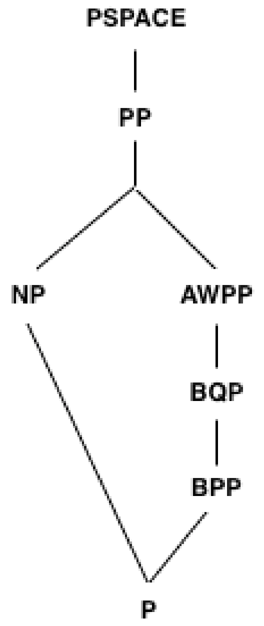


Fig. A.1: Relation between certain computational complexity classes. A bold line between two classes indicates the lower class is contained within the higher.

Definition A.3 (**BPP**). *A language $\mathcal{L}$ is in **BPP** if and only if there exists a polynomial-time probabilistic Turing machine M, such that*

For every $x \in \mathcal{L}$, M accepts with probability at least $2/3$

For every $x \notin \mathcal{L}$, M accepts with probability at most $1/3$

Definition A.4 (**PP**). *A language $\mathcal{L}$ is in **PP** if and only if there exists a polynomial-time probabilistic Turing machine M, such that*

For every $x \in \mathcal{L}$, M accepts with probability strictly greater than $1/2$

For every $x \notin \mathcal{L}$, M accepts with probability strictly less than $1/2$

Definition A.5 (**PSPACE**). *A language $\mathcal{L}$ is in **PSPACE** if and only if there exists a deterministic Turing machine M working in polynomial space (see [159]), such that*

1. *For every $x \in \mathcal{L}$, M accepts*

2. *For every $x \notin \mathcal{L}$, M rejects*

Appendix B

Proofs and advice: the classical case

The study of non-uniform classical computation begins with polynomial-sized Boolean circuits. These circuits can equivalently be viewed as Turing Machines that take polynomial-sized advice bit-strings. These strings only depend on the size of the input and not the input itself. If the string were to depend on every input then we could just encode the solution to any problem for that input and be able to decide any language. The class of decision problems that are solved by a (uniform) deterministic classical computer with classical advice is denoted **P/poly**, where the suffix **/poly** denotes a classical advice bit-string.

Definition B.1. **P/poly** *is the class of languages* $\mathcal{L} \subseteq \{0,1\}^n$ *for which there exists a poly-time uniform classical circuit family* $\{\mathcal{C}_x\}$ *and a set of bit-strings* $\{y_n\}_{n\geq 1}$ *of length* $d(n)$ *for some polynomial* d*, such that for all strings* $x \in \{0,1\}^n$*,* $x \in \mathcal{L}$ *if and only if* $\mathcal{C}_x$ *accepts for* (x, y_n) *as input.*

Since we will be considering probabilistic processes in full generality, it is worth defining the relevant class of computation with advice where processes are not deterministic. In full generality, we allow the possibility that the advice bit-strings are sampled from a probability distribution for each input size – we denote such advice as "randomized advice" denoted by the suffix **/rpoly**. In addition to this, we allow the uniform circuits to accept inputs with some error as is normal in efficient probabilistic computation (cf. the definition of **BGP**). Therefore the class **BPP/rpoly** of problems solved (with some error) by a (uniform) classical circuit with randomized advice can now be defined.

Definition B.2. **BPP/rpoly** *is the class of languages* $\mathcal{L} \subseteq \{0,1\}^n$ *for which there exists a poly-time uniform classical circuit family* $\{\mathcal{C}_x\}$ *and a set of randomized advice* *bit-strings* $\{y_n\}_{n\geq 1}$ *of length* $d(n)$ *for some polynomial* d*, such that for all strings* $x \in \{0,1\}^n$*:*

1. *If* $x \in \mathcal{L}$ *then* $\mathcal{C}_x$ *accepts with probability at least* $2/3$ *given* (x, y_n) *as input.*

2. *If* $x \notin \mathcal{L}$ *then* $\mathcal{C}_x$ *accepts with probability at most* $1/3$ *given* (x, y_n)*.*

Interestingly, despite the ability to use probabilistic processes, via derandomisation arguments it can be shown that $\mathbf{BPP/rpoly} = \mathbf{P/poly}$ [105, 116].

In the case where an efficient computer is given a proof from some untrusted provider we have already mentioned the classical complexity class **NP** but this is not the most general class for probabilistic computation. If the efficient classical computer accepts some input with some error, then this is in the remit of Merlin-Arthur games with complexity as follows.

Definition B.3. **MA** *is the class of languages* $\mathcal{L} \subseteq \{0,1\}^n$ *for which there exists a poly-time uniform classical circuit* $\{\mathcal{C}_x\}$ *and a polynomial* d*, such that for all strings* $x \in \{0,1\}^n$*:*

1. *If* $x \in \mathcal{L}$ *then there exists a proof* $z \in \{0,1\}^{d(n)}$ *such that* $\mathcal{C}_x$ *accepts with probability at least* $2/3$ *given* (x, z) *as input.*

2. *If* $x \notin \mathcal{L}$ *then* $\mathcal{C}_x$ *accepts with probability at most* $1/3$ *given* (x, z) *as input, for all proofs* z*.*

The existential quantifiers in the above definition rigorously capture the notion of a circuit having to 'verify' the proof. It immediately follows that $\mathbf{NP} \subseteq \mathbf{MA}$. This definition will also allow us to readily present the quantum analogue to this class along with its analogue for all possible general theories.

Appendix C

Proofs and advice: the quantum case

The class of decision problems that can be solved by an efficient quantum computer with quantum advice, denoted by **BQP/qpoly**, is defined as follows.

Definition C.1. BQP/qpoly *is the set of languages* $\mathcal{L} \subseteq \{0,1\}^n$ *for which there exists a poly-time uniform quantum circuit family* $\{\mathcal{Q}_x\}$ *and a set of (possibly non-uniform) states* $\{|\psi_n\rangle\}_{n\geq 1}$ *of* $d(n)$ *qubits for some polynomial* d*, such that for all strings* $x \in \{0,1\}^n$*:*

1. *If* $x \in \mathcal{L}$ *then* $\mathcal{Q}_x$ *accepts with probability at least* $2/3$ *given* $|x\rangle|\psi_n\rangle$ *as input.*

2. *If* $x \notin \mathcal{L}$ *then* $\mathcal{Q}_x$ *accepts with probability at most* $1/3$ *given* $|x\rangle|\psi_n\rangle$*.*

The class of decision problems for which a "yes" outcome can be verified in quantum poly-time, with help from a poly-size quantum proof, or witness, state, denoted **QMA**, is defined as follows.

Definition C.2. QMA *is the set of languages* $\mathcal{L} \subseteq \{0,1\}^n$ *for which there exists a poly-time uniform quantum circuit* $\{\mathcal{Q}_x\}$ *and a polynomial* d*, such that for all strings* $x \in \{0,1\}^n$*:*

1. *If* $x \in \mathcal{L}$ *then there exists a* $d(n)$*-qubit quantum proof* $|\phi\rangle$ *such that* $\mathcal{Q}_x$ *accepts with probability at least* $2/3$ *given* $|x\rangle|\phi\rangle$ *as input.*

2. *If* $x \notin \mathcal{L}$ *then* $\mathcal{Q}_x$ *accepts with probability at most* $1/3$ *given* $|x\rangle|\phi\rangle$ *as input, for all proofs* $|\phi\rangle$*.*

The existential quantifiers in the above definition of **QMA** rigorously capture the notion of a quantum circuit having to 'verify' the quantum proof.

Appendix D

Comparing definitions of higher-order interference

The original definition of higher-order interference was in the framework of quantum measure theory [124, 125]. The definition revolves around the concepts of 'histories' and the 'quantum measure'. Histories correspond to paths through space-time, a set of histories A is any collection of these paths. We will be concerned with sets of histories with some initial condition s and some final condition e along with an intermediate condition i that 'the history passes through slit i'. We label these sets of histories A_i^{se}. In an interference experiment it is common to have some way to create a path difference between the different slits, either by introducing some 'phase shifter' or by moving the final detection point, we label this data t.

The quantum measure μ associates some probability to each set of histories, which should be thought of as the probability that a particle 'has a history from that set'. In general μ will depend on the experimental control t.

The existence of higher-order interference in this framework is as follows:

Definition D.1. *n-th order interference* $\iff \exists s, e$ *s.t.*

$$\mu\left[\bigcup_{i\in\mathcal{P}} A_i^{se}\right](t) \neq \sum_{I\subset\mathcal{P}} (-1)^{n-|I|+1} \mu\left[\bigcup_{i\in I} A_i^{se}\right](t)$$

Given this definition, we provide a translation to the operational definition being employed in this thesis:

	QMT	GPT
Initial condition	s	$\|s)$
Final condition	e	$(e\|$
Experimental control	t	T
Probability of path i	$\mu[A_i^{se}](t)$	$\mathcal{C}_{se_{\{i\}}}(T)$
Probability of subset I	$\mu[\bigcup_{i\in I} A_i^{se}](t)$	$\mathcal{C}_{se_I}(T)$

Note that some ambiguity is introduced in switching to the operational framework, which $e_I \in \mathcal{E}_I$ should be picked? Other approaches to defining higher-order interference in operational theories [72] have required sufficient structure to define a set of 'filters', $\{F_I\}$, for the theory. These are transformations that represent the action of leaving open some subset of slits I whilst closing the others, in which case $(e_I| = (e|F_I$. However, arbitrary theories do not have sufficient structure to define filters and so one must consider all possible choices $(e_I|$ with the correct support. This leads to the following definition of nth-order interference,

Definition D.2. *n-th order interference* $\iff \exists s, e$ *s.t.*

$$\mathcal{C}_{s,e}(T) \neq \sum_{I \subset \mathcal{P}} (-1)^{n-|I|+1} \mathcal{C}_{s,e_I}(T),$$

$\forall (e_I| \in \mathcal{E}_I$.

The introduction of the '$\forall$' statement compared to the original definition is due to the ambiguity in choosing which effect corresponds to blocking some subset of paths. In the main text the explicit dependence on T in the above equation will be suppressed, as $\mathcal{C}_{se}$ has already been defined as a function from the phase group to probabilities.

For example, the existence of second-order interference implies that there exists $|s)$ and $(e|$ such that

$$\mathcal{C}_{s,e} \neq \mathcal{C}_{s,e_{\{0\}}} + \mathcal{C}_{s,e_{\{1\}}},$$

$\forall\ |e_{\{i\}}) \in \mathcal{E}_i$. Whilst the existence of third-order interference corresponds to the existence of, $|s)$ and $(e|$ such that

$$\mathcal{C}_{s,e} \neq \mathcal{C}_{s,e_{\{0,1\}}} + \mathcal{C}_{s,e_{\{1,2\}}} + \mathcal{C}_{s,e_{\{2,0\}}} - \mathcal{C}_{s,e_{\{0\}}} - \mathcal{C}_{s,e_{\{1\}}} - \mathcal{C}_{s,e_{\{2\}}},$$

$\forall\ |e_I) \in \mathcal{E}_I$.

We consider the above for the case of quantum theory to provide some intuition for the definitions. Firstly, we show the existence of second-order interference. Define our paths by $p_i := (|i\rangle\langle i|, |i\rangle\langle i|)$, then choose $|s) = |+\rangle\langle +| = (e|$. Then $(e_{\{i\}}| \in \{r_i |i\rangle\langle i|\}$ where r_i is an arbitrary positive real number. The phase group is given by $\mathcal{P} := \{e^{i\theta_0}|0\rangle\langle 0| + e^{i\theta_1}|1\rangle\langle 1|\}$. It is then simple to show that,

$$\mathcal{C}_{s,e}(T) = \cos^2\left(\frac{\theta_0 - \theta_1}{2}\right),$$

whilst,

$$\mathcal{C}_{s,e_{\{0\}}}(T) + \mathcal{C}_{s,e_{\{1\}}}(T) = \frac{r_0 + r_1}{\sqrt{2}}.$$

It is then simple to see that, as functions of θ_i,

$$\cos^2\left(\frac{\theta_0-\theta_1}{2}\right)\neq\frac{r_0+r_1}{\sqrt{2}},$$

for any choice of r_i, i.e. $(e_{\{i\}}|$. Therefore—by our definition—quantum theory has second-order interference as we would expect.

Next we consider our definition of third-order interference for quantum theory. We consider a specific choice of $|s)$ and $(e|$, and note that this can be simply—but tediously—generalised to all choices. Consider $|s)=\frac{1}{3}(|0\rangle+|1\rangle+|2\rangle)(\langle 0|+\langle 1|+\langle 2|)=(e|$, and the phase group, $\mathcal{P}=\{e^{i\theta_0}|0\rangle\langle 0|+e^{i\theta_1}|1\rangle\langle 1|+e^{i\theta_2}|2\rangle\langle 2|\}$. Then let $(e_{\{i,j\}}|=\frac{1}{3}(|i\rangle+|j\rangle)(\langle i|+\langle j|)$ and $(e_{\{i\}}|=\frac{1}{3}|i\rangle\langle i|$. Note that these are sub-normalised effects. It is then simple to check our definition for this particular choice of $|s)$ and $(e|$. The interference patterns can be written as,

1.
$$\begin{aligned}\mathcal{C}_{s,e}(T)&=\frac{1}{9}|e^{i\theta_0}+e^{i\theta_1}+e^{i\theta_2}|^2\\&=\frac{1}{9}\left(3+\sum_{i>j}e^{i(\theta_i-\theta_j)}+e^{i(\theta_j-\theta_i)}\right),\end{aligned}$$

2.
$$\begin{aligned}\mathcal{C}_{s,e_{\{i,j\}}}(T)&=\frac{1}{9}|e^{i\theta_i}+e^{i\theta_j}|^2\\&=\frac{1}{9}\left(2+e^{i(\theta_i-\theta_j)}+e^{i(\theta_j-\theta_i)}\right),\end{aligned}$$

3.
$$\mathcal{C}_{s,e_{\{i\}}}(T)=\frac{1}{9}|e^{i\theta_i}|^2=\frac{1}{9}$$

and so,

$$\mathcal{C}_{s,e}(T)=\sum_{i>j}\mathcal{C}_{s,e_{\{i,j\}}}(T)-\sum_{i}\mathcal{C}_{s,e_{\{i\}}}(T).$$

This proves that the particular choice of state $|s)$ and effect $(e|$ do not give higher-order interference for quantum theory. This can, however, be readily generalised to hold for any choice, and so demonstrates that quantum theory does not exhibit higher-order interference.

Bibliography

[1] M. A. Nielsen and I. L. Chuang, *Quantum computation and Quantum information*, Cambridge University press, 2000.

[2] A. Einstein, *Letter to Max Born, 3rd March 1947, In The Born-Einstein Letters; Correspondence between Albert Einstein and Max and Hedwig Born from 1916 to 1955,* Walker, New York, 1971

[3] R. Feynman, R. Leighton and M. Sands, *The Feynman Lectures on Physics,* Ch. 37, Vol. 1, 1963, Basic Books.

[4] D. Deutsch, *Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer,* Proceedings of the Royal Society of London A, Volume 400, Issue 1818 (1985).

[5] W. van Dam, *Implausible consequences of superstrong nonlocality*, arXiv:quant-ph/0501159, 2005.

[6] B. S. Cirel'son, *Quantum generalizations of Bell's inequality,* Lett. Math. Phys. 4:2, 93-100 (1980).

[7] A. J. Short and J. Barrett, *Strong nonlocality: A trade-off between states and measurements*, New Journal of Physics 12, 033034, 2010.

[8] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, *Teleportation in General Probabilistic theories*, arXiv:quant-ph/0805.3553, 2008.

[9] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, *A generalized no-broadcasting theorem,* Phys. Rev. Lett 99.240501, 2007.

[10] J. Barrett, L. Hardy, and A. Kent, *No Signalling and Quantum key Distribution,* Phys. Rev. Lett 95, 010503, 2005.

[11] D. Gross, M. Mueller, R. Colbeck, and O. Dahlsten, *All reversible dynamics in maximal non-local theories are trivial*, Phys. Rev. Lett. 104, 080402, 2010.

[12] D. S. Abrams and S. Lloyd, *Nonlinear quantum mechanics implies polynomial-time solution for NP-complete problems*, Phys. Rev. Lett 81, 3992-3995, 1998.

[13] D. Bacon, *Quantum computational complexity in the presence of closed timelike curves*, Phys. Rev. A 70, 032309, 2004.

[14] S. Aaronson, *Quantum Computing and Hidden Variables II: The Complexity of Sampling Histories*, arXiv:quant-ph/0408119, 2004.

[15] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Probabilistic theories with purification*, Phys. Rev. A 81, 062348, 2010.

[16] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Informational derivation of Quantum Theory*, Phys. Rev. A 84, 012311, 2011.

[17] J. Barrett, *Information processing in generalised probabilistic theories*, Phys. Rev. A 75 No. 3, 032304, 2007.

[18] L. Hardy, *Quantum theory from five reasonable axioms*, arXiv:quant-ph/0101012v4, 2001.

[19] L. Hardy, *Reformulating and reconstructing quantum theory*, arXiv:quant-ph/1104.2066v3, 2011.

[20] G. de la Torre, L. Masanes, A. J. Short and M. P. Mueller, *Deriving quantum theory from its local structure and reversibility,* Phys. Rev. Lett. 109, 090403 (2012).

[21] L. Masanes and M. P. Mueller, *A derivation of quantum theory from physical requirements,* New J. Phys. 13:063001, 2011.

[22] L. Masanes, M. P. Mueller R. Augusiak and D. Perez-Garcia, *Existence of an information unit as a postulate of quantum theory,* PNAS, vol 110, no 41, page 16373, 2013.

[23] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Quantum from principles,* Chapter in "Quantum Theory: Informational Foundations and Foils", G. Chiribella and R. Spekkens eds., Springer (2016).

[24] A. Peres, *Quantum Theory: Concepts and Methods,* Kluwer Academic, Boston, 1995.

[25] H. R. Brown, *Michelson, FitzGerald and Lorentz: the origins of relativity revisited,* URL: http://philsci-archive.pitt.edu/987, 2003.

[26] I. M. Gelfand and M. A. Naimark, *On the embedding of normed rings into the ring of operators in Hilbert space,* Matematiceskij sbornik 54 (2) (1943) 197-217.

[27] I. E. Segal, *Irreducible representations of operator algebras,* Bulletin of the American Mathematical Society 53 (2) (1947) 73-88. doi:10.1090/S0002-9904-1947-08742-5.

[28] J. Henson, R. Lal, and M. Pusey, *Theory-independent limits on correlations from generalised Bayesian networks*, New J. Phys. 16 113043, 2014.

[29] S. Popescu and D. Rohrlich, *Quantum nonlocality as an axiom*, Found. Phys. Volume 24, Issue 3, pp379-385, 1994.

[30] W. K. Wootters, *Local accessibility of quantum states,* Complexity, entropy and the physics of information, 8 (1990) 39-46.

[31] L. Hardy and W. K. Wootters, *Limited holism and real-vector-space quantum theory,* Foundations of Physics 42 (3) (2012) 454-473.

[32] S. Massar, S. Pironio and D. Pitalúa-García, *Hyperdense coding and superadditivity of classical capacities in hypersphere theories,* New J. Phys. 17, 113002 (2015).

[33] P. Janotta, C. Gogolin, J. Barrett and N. Brunner, *Limits on non-local correlations from the structure of the local state space,* New J. Phys. 13 (2011) 063024.

[34] S. Massar and M. K. Patra, *Information and communication in polygon theories,* Phys. Rev. A 89, 052124 (2014).

[35] G. D'Ariano, F. Manessi, and P. Perinotti, *Determinism without causality,* arXiv:quant-ph/1301.7578, 2013.

[36] D. Aharonov, A. Kitaev, and N. Nisan, *Quantum circuits with mixed states,* arXiv:quant-ph/9806029, 1998.

[37] G. Kuperberg, *How hard is it to approximate the Jones polynomial?* arXiv:quant-ph/0908.0512v2, 2014.

[38] R. Jozsa, D. Shepherd, and M. Bremner, *Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, rspa20100301, 2010.

[39] J. Fitzsimons, T. Morimae, and K. Fujii, *On the hardness of classically simulating the one clean qubit model*, Phys. Rev. Lett. 112, 130502, 2014.

[40] A. Ambainis, L. Schuman, and U. Vazirani, *Computing with highly mixed states*, arXiv:quant-ph/0003136v1, 2000.

[41] L. Hemaspaandra and M. Ogihara, *The complexity theory companion*, Springer, 2002.

[42] H. Barnum, *Private correspondence*, 2014.

[43] J. Machta, *Phase information in quantum oracle computing,* arXiv:quant-ph/9805022v1, 1998.

[44] C. Bennett, E. Bernstein, G. Brassard, and U. Vazirani, *Strengths and weaknesses of quantum computing*, SIAM J. Comput., 26(5), 15101523, 1997.

[45] L. Fortnow and J. Rogers, *Complexity limitations on quantum computation*, Journal of Computer and System Sciences 59, 240-252, 1999.

[46] J. Kobler, U. Schoning, and J. Toran, *Graph isomorphism is low for PP,* Computational Complexity 2.4 (1992): 301-330.

[47] L. Valiant and V. Vazirani, *NP is as easy as detecting unique solutions,* Theoretical Computer Science 47: 85-93, 1986.

[48] A. Bouland, L. Mančinska and X. Zhang *Complexity classification of two-qubit commuting Hamiltonians,* arXiv:1602.04145, 2016.

[49] M. Mueller and C. Ududec, *The structure of reversible computation determines the self-duality of quantum theory*, Phy. Rev. Lett. 108, 130401, 2012.

[50] L. Hardy, *Probability theories with dynamical causal structure: A new framework for quantum gravity*, arXiv:gr-qc/0509120v1, 2005.

[51] L. Hardy, *Quantum gravity computers: On the theory of computation with indefinite causal structure*, arXiv:quant-ph/0701019v1, 2007.

[52] J. Watrous, *Quantum computational complexity*, arXiv:quant-ph/0804.3401, 2008.

[53] S. Fenner, *PP-lowness and simple definition of AWPP*, Theory of Computing Systems, Volume 36, Issue 2, 2003.

[54] S. Fenner, L. Fortnow, and S. Kurtz,*Gap-definable counting classes,* Journal of Computer and System Sciences Volume 48, Issue 1, February 1994, Pages 116-148.

[55] S. Fenner, L. Fortnow, S. Kurtz, and L. Li, *An oracle builders toolkit*, Proceedings of the 8th IEEE structure in complexity theory conference, 1993.

[56] L. Li, *On the counting functions*, PhD thesis, University of Chicago, 1993.

[57] A. Yao, *Separating the polynomial time hierarchy by oracles: part 1*, Proc. 26th IEEE FOCS, 1985.

[58] N. de Beaudrap, *On computation with 'probabilities' modulo k*, arXiv:cs.CC/1405.7381v2.

[59] J. Allen, *Treating time travel quantum mechanically*, Phys. Rev. A 90(4) 042107, 2014.

[60] M. A. Nielsen, *Computable functions, quantum measurements, and quantum dynamics*, Phys.Rev.Lett. 79 (1997) 2915-2918.

[61] G. Chiribella, G. D'Ariano, P. Perinotti and B. Valiron, *Quantum computations without definite causal structure*, Phys. Rev. A 88, 022318, 2013.

[62] M. Araújo, F. Costa and C. Brukner *Computational advantage from quantum-controlled ordering of gates*, Phys. Rev. Lett. 113, 250402, 2014.

[63] S. Aaronson and A. Arkhipov, *The Computational Complexity of Linear Optics*, Proc. of the Forty-third Annual ACM Symposium on Theory of Computing (STOC 2011), pp. 333-342, 2011.

[64] P. Shor, *Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer*, SIAM J. Sci. Statist. Comput. 26, 1484, 1997.

[65] S. Aaronson, *Quantum computing, postselection, and probabilistic polynomial-time*, Proc. R. Soc. A, 461, 3473-3482, 2005.

[66] C. M. Lee and M. J. Hoban, *Bounds on the power of proofs and advice in general physical theories*, Proc. R. Soc. A 472 (2190), 20160076 (2016).

[67] C. M. Lee and M. J. Hoban, *The Information Content of Systems in General Physical Theories*, EPTCS 214, 2016, pp. 22-28, 2016.

[68] S. W. Al-Safi and A. J. Short, *Simulating all non-signalling correlations via classical or quantum theory with negative probabilities*, Phys. Rev. Lett. 111, 170403, 2013.

[69] G. Oas, J. Acacio de Barros and C. Carvalhaes, *Exploring non-signalling polytopes with negative probability*, Phys. Scr. 014034, 2014.

[70] H. Pashayan, J. J. Wallman and S. D. Bartlett, *Estimating outcome probabilities of quantum circuits using quasiprobabilities*, arXiv:quant-ph/1503.07525, 2015.

[71] S. Popescu, *Nonlocality beyond quantum mechanics*, Nature Physics 10, 264-270, 2014.

[72] H. Barnum, M. P. Mueller and C. Ududec, *Higher-order interference and single system postulates for quantum theory*, New Journal of Physics 16, 123029, 2014.

[73] G. Niestegge, *Conditional probability, three-slit experiments and the Jordan structure of quantum mechanics*, Advances in Mathematical Physics 156573, 2012.

[74] J. Henson, *Bounding quantum contextuality with lack of third-order interference*, Phys. Rev. Lett. 114, 220403, 2015.

[75] M. Navascués, Y. Guryanove, M. J. Hoban and A. Acín, *Almost quantum correlations*, Nature Communications 6, 6288, 2015.

[76] T. Fritz, A. B. Sainz, R. Augusiak, J. Bohr Brask, R. Chaves, A. Leverrier and A. Acín, *Local orthogonality as a multipartite principle for quantum correlations*, Nature Communications 4, 2263, 2013.

[77] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter and M. Zukowski, *Information causality as a physical principle*, Nature 461, 1101, 2009.

[78] C. Ududec, H. Barnum and J. Emerson, *Three slit experiments and the structure of quantum theory*, Foundations of Physics, 41, 3, pages 396–405, 2011

[79] A. Yao, *Quantum circuit complexity.* FOCS, 1993

[80] M. Howard, J. Wallman, V. Veitch and J. Emerson, *Contextuality supplies the magic for quantum computation*, Nature 510, 351 (2014).

[81] G. Vidal, *Efficient classical simulation of slightly entangled quantum computations*, Phys. Rev. Lett. 91, 147902 (2003).

[82] M. Hoban, J. Wallman and D. Browne, *Generalised Bell inequality experiments and computation*, Phys. Rev. A 84, 062107 (2011).

[83] A. Datta, A. Shaji and C. Caves, *Discord and the power of one qubit*, Phys. Rev. Lett. 100, 050502 (2008).

[84] D. Stahlke, *Quantum interference as a resource for quantum speed-up*, Phys. Rev. A 90, 022302, 2014.

[85] V. Vedral, *The elusive source of quantum speed-up*, Found. Physics, 40, 8, (2010).

[86] A. Brodutch, *Discord and quantum computational resources,* Phys. Rev. A 88, 022307 (2013).

[87] M. Van den Nest, *Universal Quantum Computation with Little Entanglement,* Phys. Rev. Lett. 110, 060504 (2013).

[88] C. M. Lee and J. H. Selby, *Higher-order interference in extensions of quantum theory,* Foundations of Physics (2016), DOI:10.1007/s10701-016-0045-4, arXiv:1510.03860.

[89] C. M. Lee and J. H. Selby, *Generalised phase kick-back: the structure of computational algorithms from physical principles,* New J. Phys. 18 (2016) 033023.

[90] C. M. Lee and J. H. Selby, *Deriving Grover's lower bound from simple physical principles,* New J. Phys. 18 (2016) 093047.

[91] J. Savage, *Computational work and time on finite functions*, Journal of ACM, 17, 1972.

[92] C. Schnorr, *The network complexity and Turing machine complexity of finite functions,* Acta Informatica, 7, 1976.

[93] C. M. Lee and J. Barrett, *Computation in generalised probabilistic theories*, New Journal of Physics 17, 083001, 2015.

[94] J. Barrett, N. de Beaudrap, M. J. Hoban and C. M. Lee, *The computational landscape of general physical theories*, Forthcoming, 2016.

[95] S. W. Al-Safi and A. J. Short, *Reversible dynamics in strongly non-local Boxworld systems*, J. Phys. A: Math. Theor. 47, 325303, 2014.

[96] S. W. Al-Safi and J. Richens, *Reversibility and the structure of the local state space*, arXiv:1508.03491, 2015.

[97] S. Bravyi, D. P. DiVincenzo, R. I. Oliveira and B. M. Terhal, *The Complexity of Stoquastic Local Hamiltonian Problems*, Quant. Inf. Comp. 8 (5), 361-385, 2008.

[98] M. J. Hoban and D. E. Browne, *Stronger Quantum Correlations with Loophole-free Post-selection*, Phys. Rev. Lett. 107, 120402, 2011.

[99] M. N. Vyalyi, *$QMA = PP$ implies that PP contains PH*, Electronic Colloquium on Computational Complexity, Report No. 21, 2003.

[100] M. Mueller and L. Masanes, *Three-dimensionality of space and the quantum bit: an information-theoretic approach*, New Journal of Physics 15, 053040, 2013.

[101] H. Nishimura and T. Yamakami. *Polynomial time quantum computation with advice*, Electronic Colloquium on Computational Complexity, TR03-059, 2003.

[102] S. Aaronson, *Limitations of Quantum Advice and One-Way Communication*, Theory of Computing 1, pp. 1-28, 2005.

[103] S. Aaronson and A. Drucker, *A full characterisation of quantum advice*, Proc. of the Forty-second Annual ACM Symposium on Theory of Computing (STOC 2010), pp. 131-140, 2010.

[104] D. Aharonov, I. Arad and T. Vidick, *The Quantum PCP Conjecture*, ACM SIGACT Vol. 44 (2), pp. 47-79, 2013.

[105] L. Adleman, *Two theorems on random polynomial time*, Proc. of FOCS 78, 1978.

[106] C. Marriott and J. Watrous, *Quantum Merlin-Arthur games*, Computational Complexity 14 (2), pp. 122-152 (2005).

[107] A. Kitaev and J. Watrous, *Parallelization, amplification, and exponential time simulation of quantum interactive proof system*, Proc. of the Thirty-second Annual ACM Symposium on Theory of Computing (STOC 2000), pp. 608-617, 2000.

[108] J. Barrett and S. Pironio, *Popescu-Rohrlich correlations as a unit of nonlocality*, Phys. Rev. Lett. 95, 140401, 2005.

[109] R. M. Karp and R. J. Lipton, *Turing machines that take advice*, Enseign. Math., 28:191-201, 1982.

[110] M. Ben-Or, S. Goldwasser, J. Kilian and A. Wigderson, *Multi-prover interactive proofs: how to remove intractability*, Proc. of the Twentieth Annual ACM Symposium on Theory of Computing (STOC 88), pp. 113-131, 1988.

[111] T. Ito and T. Vidick, *A Multi-prover Interactive Proof for NEXP Sound against Entangled Provers*, Proc. of FOCS 2012, pp. 243-252, 2012.

[112] Y. T. Kalai, R. Raz and R. D. Rothblum, *How to delegate computations: The power of no-signaling proofs*, Proc. of the Forty-sixth Annual ACM Symposium on Theory of Computing, 2014.

[113] C. Pfister and S. Wehner, *If no information gain implies no disturbance, then any discrete physical theory is classical*, Nature Communications 4, 1851, 2013.

[114] M. Mueller, J. Oppenheim and O. Dahlsten, *The black hole information problem beyond quantum theory*, Journal of High Energy Physics 2012, Issue 9, 116 (2012).

[115] A. Drucker and R. de Wolf, *Quantum Proofs for Classical Theorems*, Theory of Computing, Graduate surveys 2, 2011.

[116] S. Aaronson, *QMA/qpoly Is Contained In PSPACE/poly: De-Merlinizing Quantum Protocols*, Proceedings of IEEE Complexity 2006, pages 261-273.

[117] A. Winter, *Coding theorem and strong converse for quantum channels,* IEEE Trans. Inform. Theory, vol. IT-45, 1999.

[118] T. Ogawa and H. Nagaoka, *A New Proof of the Channel Coding Theorem via Hypothesis Testing in Quantum Information Theory,* arXiv:quant-ph/0208139, 2002.

[119] R. Raz, *Exponential separation of quantum and classical communication complexity,* Proc. of 31st Annual ACM Symposium on the Theory of Computing, 358 (1999).

[120] H. Buhrman, R. Cleve and W. van Dam *Quantum Entanglement and Communication Complexity,* SIAM J.Comput. 30 (2001) 1829-1841.

[121] B. Dakić, T. Paterek and Č. Brukner, *Density cubes and higher-order interference theories*, New J. Phys. 16, 023028 (2014)

[122] K̇. Zyczkowski, *Quartic quantum theory: an extension of the standard quantum mechanics*, J. Phys. A 41, 355302-23, 2008.

[123] R. W. Spekkens, *Evidence for the epistemic view of quantum states: A toy theory*, Physical Review A, 75, 3, 032110, 2007.

[124] R. Sorkin, *Quantum mechanics as quantum measure theory*, Modern Physics Letters A, 9, 33, 3119–3127, 1994.

[125] R. Sorkin, *Quantum measure theory and its interpretation*, arXiv preprint gr-qc/9507057, 1995.

[126] P. Janotta and R. Lal, *Generalized probabilistic theories without the no-restriction hypothesis*, Physical Review A 87.5, 052131, 2013.

[127] F. Dowker, J. Henson and P. Wallden, *A histories perspective on characterizing quantum non-locality*, New Journal of Physics, 16, 3, 033033, 2014.

[128] G. Niestegge, *Three-slit experiments and quantum nonlocality*, Foundations of Physics, 43,6,805–812,2013.

[129] C. Ududec, *Perspectives on the formalism of quantum theory*, University of Waterloo, 2012.

[130] A. Garner, O. Dahlsten, Y. Nakata, M. Murao and V. Vedral, *Phase and interference in generalised probabilistic theories*, New J. Phys. 15 093044 2013.

[131] U. Sinha, C. Couteau, Z. Medendorp, I. Söllner, R. Laflamme and R. Sorkin, *Testing Born's rule in quantum mechanics with a triple slit experiment*, arXiv preprint arXiv:0811.2068, 2008.

[132] D. Park, O. Moussa and R. Laflamme, *Three path interference using nuclear magnetic resonance: a test of the consistency of Born's rule*, New Journal of Physics, 14, 11, 113025, 2012.

[133] G. Barssard, P. Hoyer and P. Tapp, *Quantum algorithms for the Collision Problem* , arXiv preprint, arXiv:quant-ph/9705002v1, 1997.

[134] D. Deutsch and R. Jozsa, *Rapid solutions of problems by quantum computation*, Proceedings of the Royal Society of London A 439: 553, 1992.

[135] R. Cleve, A. Ekert, C. Macchiavello and M. Mosca, *Quantum algorithms revisited*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, volume 454, 1998.

[136] O. Dahlsten, A. Garner, J. Thompson, M. Gu and V. Vedral, *Particle exchange in post-quantum theories*, arXiv:quant-ph/1307.2529-v2, 2013.

[137] G. N. Afanasiev, *Quantum mechanics of toroidal anyons*, J. Phys. A: Math. Gen. 24 2517, 1991.

[138] B. Zwiebach, *A first course in String Theory*, Cambridge University press, 2009.

[139] K. Korzekwa, M. Lostaglio, J. Oppenheim and D. Jennings, *The extraction of work from quantum coherence*, arXiv:quant-ph/1506.07875, 2015.

[140] B. Coecke, *Quantum picturalism*, Contemporary physics, 51, 1, 2010.

[141] S. Abramsky and B. Coecke, *A categorical semantics of quantum protocols*, Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science, 2004.

[142] H. Barnum, J. Barrett, M. Krumm and M. Mueller, *Entropy, majorization and thermodynamics in general probabilistic theories*, arXiv:quant-ph/1508.03107, 2015.

[143] G. Chiribella and C.M. Scandolo, *Entanglement and thermodynamics in general probabilistic theories*, arXiv:quant-ph/1504.07045, 2015.

[144] G. Chiribella and C.M. Scandolo, *Operational axioms for state diagonalization*, arXiv:quant-ph/1506.00380,2015.

[145] H. Barnum, M. Graydon and A. Wilce, *Some Nearly Quantum Theories*, arXiv:1507.06278, 2015.

[146] S. Aaronson, A. Bouland, J. Fitzsimons and M. Lee, *The space just above BQP,* In Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, pages 271–280. ACM, 2016.

[147] N. Bao, A. Bouland and S. P. Jordan, *Grover search and the no-signaling principle,* arXiv preprint arXiv:1511.00657, 2015.

[148] M. Boyer, G. Brassard, P. Høyer and A. Tapp, *Tight bounds on quantum searching,* arXiv preprint quant-ph/9605034, 1996.

[149] L. K. Grover, *Quantum mechanics helps in searching for a needle in a haystack,* Physical review letters, 79(2):325, 1997.

[150] D. R. Simon, *On the power of quantum computation*, 35th Annual Symposium on: Foundations of Computer Science, 116123, 1996 Proceedings.

[151] S. Popescu and D. Rohrlich, *Causality and nonlocality as axioms for quantum mechanics*, Springer, 1998.

[152] A. Sinha, A. H. Vijay and U. Sinha, *On the superposition principle in interference experiments,* Scientific reports, 5, 2015.

[153] U. Sinha, C. Couteau, T. Jennewein, R. Laflamme and G. Weihs, *Ruling out multi-order interference in quantum mechanics,* Science, 329(5990):418–421, 2010.

[154] C. Zalka, *Grover's quantum searching algorithm is optimal,* Physical Review A, 60(4):2746, 1999.

[155] A. W. Harrow, A. Hassidim and S. Lloyd *Quantum algorithm for solving linear systems of equations,* Phys. Rev. Lett. vol. 15, no. 103, pp. 150502 (2009).

[156] A. Broadbent, J. Fitzsimons and E. Kashefi, *Universal blind quantum computation,* Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009), pp. 517-526.

[157] T. Morimae and K. Fujii *Blind quantum computation protocol in which Alice only makes measurements,* Phys. Rev. A 87, 050301(R) (2013).

[158] D. Aharonov, M. Ben-Or and E. Eban, *Interactive Proofs For Quantum Computations,* arXiv:0810.5375, 2008.

[159] C. H. Papadimitriou, *Computational complexity,* Pearson, 1993.