The Space Around BQP

by

Adam Michael Bouland

B.S., Yale University (2009)
M.A.St., University of Cambridge (2010)
M.Phil., University of Cambridge (2011)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

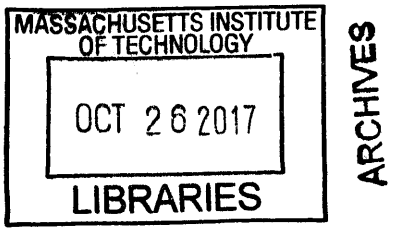
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2017

Author ..... Signature redacted .....
Department of Electrical Engineering and Computer Science
August 25, 2017

Certified by ..... Signature redacted .....
Professor Scott J. Aaronson
Visiting Associate Professor
Department of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by ..... Signature redacted .....
Professor Leslie A. Kolodziejski
Chair, Department Committee on Graduate Students
Department of Electrical Engineering and Computer Science



The Space Around BQP

by

Adam Michael Bouland

Submitted to the Department of Electrical Engineering and Computer Science
on August 25, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis explores the computational power of quantum devices from the perspective of computational complexity theory. Quantum computers hold the promise of solving many problems exponentially faster than classical computers. The computational power of universal quantum devices is captured by the complexity class BQP, which stands for "bounded-error quantum polynomial time." We hope that quantum devices will be capable of the full power of BQP in the long term. However, quantum computers are difficult to build, so the experimental devices of the near future may be incapable of universal quantum computation. As a result, a number of recent works have studied "weak" models of quantum computers which lie "below BQP."

The first part of this thesis examines the space "below BQP" and describes a number of sub-universal models of quantum computation which can nevertheless perform sampling tasks which are difficult for classical computers. We show that prior models maintain hardness when their set of quantum operations is restricted, and describe two new models of "weak" quantum computation which also show advantage over classical devices. A major theme in this work is that almost any weak device can perform hard sampling tasks. We find that almost any model which is not universal, but not known to be efficiently classically simulable, admits a speedup over classical computing for sampling tasks under plausible assumptions. This work can be seen as progress towards classifying the power of all restricted quantum gate sets.

On the other hand, quantum gravity theorists have considered modifying quantum mechanics to resolve the black hole information paradox. Inspired by these debates, the second part of this thesis explores the computational power of modified theories of quantum mechanics. We find that almost all modifications allow for drastically more power than BQP – i.e. these modifications lie "above BQP" – and we find that these speedups may be related to superluminal signaling in these models. Surprisingly, we find one model which is only slightly more powerful than BQP. Inspired by this model, we study and resolve an open problem in classical complexity related to the power of statistical-zero knowledge proof systems.

Thesis Supervisor: Professor Scott J. Aaronson
Title: Visiting Associate Professor
Department of Electrical Engineering and Computer Science

Acknowledgments

There are many people I would like to thank for their support, mentorship, and friendship over the years.

First and foremost, I would like to thank my advisor Scott Aaronson. Scott has been a constant source of support and encouragement during my Ph.D. His creative approach to research taught me how to ask interesting complexity-theoretic questions which are both relevant and under-studied. Scott has supported my many research visits across the globe and given me total intellectual freedom. He has never been demanding of my time, but is generous with his, his door always opens for discussions about obscure complexity classes, quantum physics, or his latest DWave blog post. I couldn't have asked for a more supportive or creative advisor. I would also like to thank the other members of my thesis committee, Aram Harrow and Ryan Williams, for their helpful comments and support.

This thesis couldn't have been written with the help of many brilliant and talented collaborators. I would like to thank my coauthors - namely Scott Aaronson, Itai Arad, Ning Bao, Aiden Chatwin-Davies, Lijie Chen, Lynn Chua, Joe Fitzsimons, Daniel Grier, Dhiraj Holden, Stephen Jordan, Dax Koh, Greg Kuperberg, Laura Mančinska, Mitchell Lee, George Loether, Saeed Mehraban, Jason Pollack, Miklos Santha, Justin Thaler, Prashant Vasudevan, Henry Yuen, Luzy Zhang, and Shengyu Zhang - as well as other collaborators with whom I have not (yet!) written papers - Mohammad Bavarian, Shalev Ben-David, Matt Coudron, Ankit Garg, Robin Kothari, Māris Ozols, Anand Natarajan, Luke Schaeffer and John Wright. I have learned so much from our discussions and collaborations over the years.

Many of these collaborations were forged during research visits around the world. In particular I would like to thank Miklos Santha for hosting me for several long visits to the Centre for Quantum Technologies in Singapore, as well as Joe Fitzsimons and Itai Arad for their time and mentorship during my visits. I would like to thank the CQT community for their friendship - in particular Aarthi Sundaram, Laura Mančinska & David Roberson, Jamie Sikora & Caitlin Cooke, Supartha Podder, Anurag Anshu, and Antonis Varvitsiotis, all of whom made Singapore my home away from home. I also thank Stephen Jordan for hosting me for a visit to QuICS at the University of Maryland in 2015, during which much of Chapter 9 was written, Ashley Montanaro and Laura Mančinska for hosting a visit to U. Bristol in 2016, and Tomoyuki Morimae for hosting me for a visit to the Tokyo Institute of Technology in 2016.

I was fortunate to mentor five exceptional undergraduate students through the MIT SPUR and UROP programs. I would like to thank my students Lynn Chua (now a PhD candidate in theoretical computer science at UC Berkeley, and co-author of [12]), Hyun-sub Hwang (now at Citadel capital), Mitchell Lee (now a PhD candidate in mathematics at Harvard, and co-author of the work in Chapter 7), Mark Velednitsky, and Lucy (Xue) Zhang (co-author of the work in Chapter 4) for three excellent summers of research. It was an honor working with them.

I would also like to thank the MIT theory group for making grad school enjoyable, and in particular my cohort - Mohammad Bavarian, Matt Coudron, Alan Guo, Ioana Ivan, Sepideh Mahabadi, Ludwig Schmidt, Aaron Sidford, Christos Tzamos, Madars Virza, Adrian Vladu, and Henry Yuen. From "no so great ideas" outings, to Madhu's and Aleksander's annual STOC/FOCS CBC trips, to the annual theory retreat, to the "do you even lift" mailing list, I will miss the TCS community at MIT. I would also like to thank the Parsons Laboratory and my wife's cohort for adopting me into their community as well.

Going back further, I would like to thank all of my mentors for laying the foundation for my success in grad school - in particular my Master's supervisor Anuj Dawar, my undergraduate supervisors Dan Spielman and Richard Easther, and my mentors for numerous summer internships - Risa Wechsler, Dave Pollard, and Mostafiz Chowdhury, the latter of whom hosted me for three summers at the U.S. Army Research Lab as a teenager. I would also like to thank several influential math/science teachers in my life - namely Roger Howe, Ken Zachmann, and Stan Arnold, as well as Lynn Collins, who volunteered her time to teach an accelerated middle school math class which ignited my interested in the subject as a child.

Finally, I would like to thank my family, namely my wife Alison, my parents Dean and Janet, and my brother Andrew. I couldn't have done this without their support and encouragement.

Contents

Chapter 1

Introduction

1.1 Introduction

Quantum computers, first dreamed of by Feynman [115] and Deutsch [100], have the promise of performing certain computational tasks - such as factoring integers [221] or simulating quantum mechanics - exponentially faster than classical computers. This has lead to a decades-long experimental effort to construct quantum computers, despite the immense difficulty. In recent years the underlying experimental technologies for realizing quantum computing have improved drastically. As a result, many research groups hope to produce working quantum devices in the next 1-5 years [193] which are capable of performing computational tasks which we cannot simulate with classical computers. If successful, these experimental devices will bolster the claim that quantum computing will be efficiently realizable in the long term - and at the same time gives credence to the view that the world is fundamentally quantum mechanical.

The quantum devices of the next 1-5 years will likely *not* be capable of what is called "universal quantum computation." In other words, they will not have the maximum power that quantum computers are capable of having, and which we hope they will have in the long term. Formally, the computational power of universal quantum computers is captured by the complexity class BQP - which stands for "Bounded-error Quantum Polynomial time." As a result, a natural theoretical challenge is to try to classify the space "below BQP" - i.e. to explore the space of quantum computing models which are weaker than universal quantum computation. If one can find evidence that these "weak" quantum computers can still perform computational tasks which are impossible for classical computers, then these weak models may be viable near-term experimental targets for demonstrating and advantage over classical computation. The first part of this thesis examines the "space below BQP" and describes a number of sub-universal models of quantum computation which can nevertheless perform difficult sampling tasks. A major theme in these works is that the ability to perform hard sampling tasks is pervasive. We find that just about any model which is not universal, but not known to be efficiently classically simulable, admits a speedup over classical computing for sampling tasks under plausible assumptions.

At the same time, the development of quantum computation and quantum information has had a broad impact on theoretical physics. The notion that physical theories not only describe physical objects and dynamics (particles, black holes, etc), but also describe a theory of *information processing* has opened a conceptual bridge between theoretical computer science and theoretical physics. This connection has been most salient in the discourse

over the black hole information paradox, in which physicists are reckoning with apparent inconsistencies between the expected behavior of matter and quantum information around black holes. To resolve some of these paradoxes, some have proposed modified theories of quantum mechanics to evade the conditions of this paradox. While others have considered if these theories are physically reasonable, a natural challenge is to determine if these theories are computationally reasonable. Inspired by these connections, the second part of this thesis explores the space "above BQP" - i.e. the power of quantum computers in the presence of modified theories of quantum mechanics. We find that almost all modifications allow for drastically more power than BQP, but surprisingly find one model which is merely just a bit more powerful than BQP. We also find surprising connections between the space above BQP and classical complexity.

1.2 Below BQP: classifying the power of "weak" quantum computers

In the near term, quantum devices will be incapable of universal quantum computation, and furthermore will be limited to a small number of qubits. Therefore a natural question is what sorts of "weak" quantum devices might still possess beyond-classical computational power. A breakthrough in this area came in 2010, when Aaronson & Arkhipov [11], and Bremner, Jozsa & Shepherd [72] introduced the notion of "sampling tasks". These authors showed that certain non-universal quantum devices can sample from some probability distributions that classical devices cannot sample from efficiently, under certain complexity-theoretic assumptions. Therefore, one might be able to demonstrate a "quantum advantage" over classical computation with these non-universal devices. Subsequently a flurry of works have identified sampling tasks which are difficult for classical devices [114, 74, 73, 195, 194, 111].

In Chapters 3-6, we work towards generalizing these results by classifying the space of intermediate quantum models in terms of their computational power. In Chapters 3 and 4, we generalize the results of [11, 72] by classifying the power of variants of their computational models in the presence of restricted gate sets. We show that their hardness results in fact can be extended to a broader class of related models [63, 66]. In Chapter 5, we introduce a new sampling model known as "Conjugated Clifford Sampling", show that it can perform difficult sampling tasks, and furthermore give a complete classification of restricted gate sets of this form. One can see these first three sections as progress towards fully classifying the space below BQP as defined via restricted gate sets. Our results make use of Lie theory and representation theory. Finally, in Chapter 6, we introduce a new intermediate model of computation based on permuting particles on a line; one can see this as an analogue of the work of Aaronson and Arkhipov in which the particles are distinguishable. We partially classify its computational power under different parameter settings [14].

1.2.1 Classifying beamsplitters

We begin studying the space below BQP by extending the work of Aaronson and Arkhipov [11] on the computational power of linear optics. This chapter is based on joint work with Scott Aaronson, which has been published in Physical Review A [63].

One proposed avenue for constructing quantum computers is through the manipulation of quantum states of light. In these systems, individual photons can be spread in superposition over many possible "modes". The photons can then be passed through "beamsplitters"

constructed of half-silvered mirrors, which cause the photons to recombine and "interfere" with one another.

In order to perform universal quantum computation with optical systems, one must get the photons to interact with one another. This is very difficult as photons normally do not interact with one another at all. Interaction can be achieved by passing the photons through specially engineered media (which can effectively have nonlinearities), or by creating adaptive optical networks which perform intermediate measurements and change the corresponding beamsplitters according to the outcomes [172]. Without interaction, it is unclear if passive linear optics are capable of universal quantum computation. Despite this limitation, Aaronson and Arkhipov showed that passive linear optics perform difficult sampling tasks. Therefore, these non-interacting linear optical circuits may be a viable candidate for a first demonstration of quantum advantage over classical computation. The computational advantage relies on the assumption that the single photons entering the optical network are indistinguishable from one another; their corresponding bosonic statistics mean that the probability distributions they sample from are related to the permanents of certain matrices, which are difficult to compute classically.

In Chapter 3, we extend [11] in a different direction. We consider the power of optics in the presence of a restricted set of beamsplitters and phase shifters. Specifically, Reck et. al [208] showed that if one can perform arbitrary beamsplitters and phase shifters, then one can efficiently perform arbitrary optical transformations. Both the KLM protocol and BosonSampling assume one has access to arbitrary beamsplitters and phase shifters in order to apply this result. Here we ask the question: if one has only a restricted set of beamsplitters available, does this change the power of linear optics? A priori, one could possibly construct sets of beamsplitters which are not capable of efficiently generating all optical transformations, and therefore create weaker models of quantum computation. Surprisingly, we answer this question in the negative by showing that any nontrivial[1] beamsplitter is itself universal for optics [63].

Therefore linear optics with restricted sets of beamsplitters are either equally as powerful as those with arbitrary beamsplitters, so long as the beamsplitter is nontrivial. This fully classifies the computational power of restricted linear optical networks. Our proof makes extensive use of representation theory and the classification of finite subgroups of $SU(3)$ [109, 143, 139] in order to show that arbitrary beamsplitters generate continuous sets of transformations.

1.2.2 Classifying commuting Hamiltonians

In Chapter 4, we consider the power of a different model of "weak" quantum computing - namely "Instantaneous Quantum Computing" or IQP [72]. This chapter is based on joint work with Laura Mančinska and Xue Zhang, which has been published in CCC'16 [66].

In general quantum gates do not commute with one another - the order in which one applies them can change the resulting operation. This is true of classical circuits as well - for instance AND and OR gates do not commute with one another. The IQP model captures the power of quantum computing when the quantum gates are required to commute with one another. Therefore the order in which they are applied is irrelevant. At first glance this seems to be very weak; it is not clear that one can perform even universal *classical* computation in this model. Despite its weakness, Bremner, Jozsa and Shepherd showed

[1] Here nontrivial simply means that it mixes modes - any beamsplitter which does not mix modes is clearly efficiently classically simulable on single-photon inputs.

that such quantum circuits can perform sampling tasks which are impossible for classical randomized algorithms [72] to simulate exactly, assuming the Polynomial Hierarchy does not collapse to the third level. They later extended this to rule out even approximate classical simulations to additive error under additional complexity assumptions [74].

In these prior works, it was shown that certain commuting circuit families - specifically those diagonal in the X basis - are capable of hard sampling tasks. In Chapter 4, we aim to *classify* the entire space of which commuting gate sets can perform difficult sampling tasks. For technical reasons, we simplify the problem further by considering a continuous version of quantum gates known as Hamiltonians. We suppose one has access to a fixed commuting two-qubit Hamiltonian H, which one can apply to arbitrary pairs of qubits for arbitrary continuous amounts of time. Our goal is to classify the computational power of this model in terms of H. Clearly if H does not generate entanglement, then it cannot be used to perform difficult sampling problems, as the model would be efficiently classically simulable. Our result states that all other H are capable of hard sampling problems:

Theorem 1.2.1. *[66] Any two-qubit Hamiltonian H which generates entanglement can be used to perform hard sampling problems, assuming the polynomial hierarchy is infinite.*

The proof of this fact is quite different from the proof in Chapter 3. Since we are considering Hamiltonians rather than discrete gates, our gate set is by definition infinite. Therefore, we can use tools from continuous mathematics - in particular Lie algebras - to address our classification problem. Our proof also made use of a result of Fefferman *et al.* [112] which closed a gap in our proof techniques for Hamiltonians of the form $X \otimes X$.

1.2.3 New model: classifying conjugated Clifford circuits

In Chapters 3 and 4, we considered existing models with hardness of sampling results, and showed that these hardness of sampling results are pervasive - i.e. *any* similar model with restricted gate sets can also perform hard sampling tasks. In Chapter 5, we introduce a new model of "weak" quantum computing with similar properties. This chapter is based on joint work with Joseph Fitzsimons and Dax Koh (forthcoming) [65].

In particular, we consider the power of Conjugated Clifford Circuits (CCCs). A Clifford circuit is a quantum circuit which begins in the computational basis, performs a discrete set of possible transformations - namely those generated by CNOT, H and P gates - and then measures in the computational basis. This model does not seem to be universal for BQP, because the set of allowed transformations is finite. In fact, Clifford circuits are efficiently classically simulable by the Gottesman-Knill Theorem [133]. Therefore Clifford circuits are a computationally weak subset of quantum computing.

An interesting fact is Gottesman and Knill's simulation algorithm for Clifford circuits breaks if the gates are conjugated by a one-qubit unitary U. We call such a circuit a "Conjugated Clifford Circuit" (CCC). This transformation manifestly keeps the set of quantum transformations allowed discrete. However there is no longer an efficient way to keep track of the quantum state as the circuit evolves. A natural question is: can CCCs perform hard sampling tasks? If so, which U allow one to perform hard sampling tasks?

Clearly if U is a Clifford element, then Clifford circuits conjugated by U can be efficiently simulated by classical computers by the Gottesman-Knill theorem, so are incapable of hard sampling tasks. A similar argument holds if is $U = CR_Z(\theta)$ - i.e. if U is a Z rotation followed by a Clifford element - since Z rotations do not affect measurement statistics in the

computational basis. In this section, we show that *any* other U allows one to perform hard sampling tasks with CCCs. This fully classifies the computational power of CCCs.

Theorem 1.2.2 ([65]). *Conjugated Clifford circuits with any $U \neq CR_Z(\theta)$ (where C is a Clifford gate) can perform hard (exact) sampling tasks, assuming the polynomial hierarchy is infinite.*

This fully classifies the complexity of exactly simulating CCCs. Our proof makes extensive use of postselection gadgets, as well as recent progress on the problem of efficient circuit compilation for general quantum circuits [212]. Furthermore, we extend this hardness result to approximately simulating CCCs, under some additional complexity-theoretic assumptions:

Theorem 1.2.3 ([65]). *Assuming a certain average-case hardness conjecture, CCCs can perform sampling tasks which are hard for classical computers, even up to constant additive error in the simulation.*

Thus CCCs may be a viable candidate for near-term demonstrations of quantum advantage.

1.2.4 New model: the power of Ball permuting

The models considered in Chapters 4 and 5 were based on restricted gate sets acting on qubits. However, many physical systems are described by Hilbert spaces which are not tensor products of qubits. In Chapter 6, we introduce a new model of intermediate quantum computing, known as the quantum ball permutations, which is based on a non-tensor product Hilbert space. This chapter is based on joint work with Scott Aaronson, Greg Kuperberg, and Saeed Mehraban, which has been published in STOC'17 [14].

Specifically, the system consists of n perfectly distinguishable balls on a line - labeled $1 \dots n$. The quantum operations allowed are partially exchanging adjacent balls in superposition. Therefore, the models is described by the Hilbert space $\mathbb{C}S_n$ - i.e. the vector space with basis consisting of all possible permutations of the n balls. This model is inspired by certain quantum field theories in 1+1 dimensions.

In this chapter, we classify the power of the ball permuting model in various regimes [14]. The power of the model depends on which input states are allowed into the system. If the input state is $|123 \dots n\rangle$, then we show that this model is very weak - in particular it is equivalent to the "one clean qubit" model of Knill and Laflamme [171]. However, this model is not known to be efficiently classically simulable, as it can perform some task (namely approximating the trace of exponentially large matrices) which seems beyond the reach of efficient classical computation. Therefore this model seems to be of intermediate power. We also show that if one allows arbitrary input states in the model, then one can recover BQP.

1.3 Above BQP: a computational lens on modifications to quantum theory

Quantum theory is one of the most precisely tested theories in human history. For instance, the fine structure constant predicted by quantum theory has been verified to many decimal places [170]. Nevertheless, there are several reasons one might consider modifying quantum theory. First, one might believe that quantum theory is merely an approximation to some

underlying theory; therefore it is reasonable to consider small perturbations to quantum mechanics. Second, several aspects of quantum theory are troublesome from the perspective of quantum foundations; by modifying the theory one might be able to reconcile quantum theory with a more deterministic universe. Third, modifications to quantum theory might help resolve current debates in quantum gravity. Recently, the firewall paradox [35] has reopened the debate over the interaction between quantum theory and general relativity in the presence of black holes. Several researchers have proposed modifications to quantum theory which may resolve the black hole information paradox [155, 181, 147].

In Chapters 7-9 of the thesis, we consider modifications of quantum theory from a computational perspective. Specifically, we consider several modifications to quantum theory, and describe their computational power from a complexity-theoretic perspective. In some cases these modifications only increase the power of quantum computation slightly. In others, the modifications drastically boost the power of quantum computation - which one may view as grounds for rejecting these theories. Along the way, we describe several unexpected connections to classical complexity theory.

1.3.1 Non-collapsing measurements

In the quantum world, measurement is a destructive process. Prior to measurement, a quantum state can be in a superposition of many possible measurement outcomes. Post measurement, the state randomly "collapses" to one of the possible outcomes. In Chapter 7, we consider modifying quantum theory by allowing for *non-collapsing* measurements. This chapter is based on joint work with Scott Aaronson, Joseph Fitzsimons, and Mitchell Lee, which has been published in ITCS'16 [13].

A *non-collapsing* measurement allows one to independently sample from the distribution on outcomes one would normally see in quantum theory - but without disturbing the underlying quantum state. We call the resulting class CQP for "Collapse-free quantum polynomial time." We likewise define a non-adaptive version of this class naCQP in which the quantum transformations cannot depend on the non-collapsing measurement outcomes. The definition of this class in inspired by Aaronson's prior work on quantum computing with hidden variable theories [2][5].

We find that this modification to quantum theory produces a class which is "just a bit" stronger than BQP [13]. In particular we show this class can break "statistical zero knowledge proof systems" in a black-box manner. In these proof systems, a prover convinces another of the veracity of a statement but reveals nothing else. This idea is extremely useful in cryptography – for example when two mistrustful parties want to prove they have performed a task, without revealing any of their private information. Breaking such cryptographic primitives in a black-box manner is known to be impossible for standard quantum computers [2]. Therefore, quantum computers equipped with non-collapsing measurement do have power beyond that of BQP.

At the same time, we show that at least the non-adaptive version of this class (naCQP) is not too powerful. In particular, we show that search in naCQP requires at least $N^{1/4}$ time

[2]Hidden variable theories are statistical interpretations of quantum mechanics in which measurement outcomes are pre-determined; the apparent randomness in quantum mechanics is merely due to the Bayesian ignorance of the experimenter. Aaronson considered the power of quantum computing, assuming a hidden variable theory is true, and that furthermore one has ability to see the evolution of the hidden variables as the computation unfolds [5]. He gave similar results to ours, but unfortunately there is a bug in the proof of his lower bound for search which we cannot resolve. Our work provides another model with similar properties for which we can prove the lower bound for search.

- and therefore such devices cannot solve NP-hard problems in a black-box manner [13]. In contrast, most known modifications to quantum theory allow search in $O(\log N)$ time [22]. Therefore this modification seems unusual in that its computational power lies "just a bit above" BQP.

1.3.2 Connections to classical complexity: SZK vs PP

Examining the computational power of non-collapsing measurements raises a number of questions in classical complexity theory. In Chapter 8, we explore a purely classical complexity theoretic question regarding the power of SZK proof systems and PP algorithms - two models of computation which are "beyond BQP" in their computational power. This chapter is based on joint work with Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan, which will be published shortly in FOCS'17 [64].

A natural problem to study is: what classical models of computation can efficiently simulate BQP? Such results provide upper bounds on the power of BQP. Clearly one can simulate BQP in exponential time by keeping track of the entire wavefunction - and therefore BQP $\subseteq$ EXP. In fact, by the Feynman sum-over-paths approach to quantum theory, one can reduce problems in BQP to questions about exponentially long sums of numbers. One can evaluate these sums in exponential time, but using merely a polynomial amount of memory. Therefore BQP $\subseteq$ PSPACE. Less obviously, one can take these ideas further to simulate quantum computations with randomized algorithms with unbounded error: this is known as the class PP [24].

In Chapter 7, we argued that CQP is merely "just a bit above" BQP in its computational power. Therefore, a natural goal is to show that CQP is contained in PP as well - this would show that both classes obey the same upper bound. One obstacle to proving this, however, was the fact that SZK is contained in CQP. Surprisingly, it is open whether or not SZK is contained in PP. Therefore, any proof that CQP $\subseteq$ PP would require answering this open question.

Therefore, in this chapter of the thesis, we consider the difficulties in placing SZK in PP. In particular, we give an oracle relative to which SZK is not contained in PP [64]. This answers an open question of Watrous from 2002 [1], and generalizes prior work of Vereshchagin [238] and Aaronson [2, 8]. It also implies that any proof that SZK $\subseteq$ PP would require nonrelativizing techniques[3]. Therefore, it may be extremely difficult to prove CQP $\subseteq$ PP. This raises the possibility that CQP may be bigger than previously considered as a complexity class.

We also describe a surprising connection between this problem and "polarization" - the process of amplifying errors in SZK. In particular, we show that our oracle separation implies that there are limits to error amplification algorithms for SZK [64].

1.3.3 Modified QM and Grover Search

Recently, high energy physics has been embroiled in debate over the firewall paradox, which arises from an apparent inconsistency between quantum mechanics and black hole physics. This is a serious obstacle to constructing a theory of quantum gravity. The firewall paradox also has fascinating connections with quantum computing theory.

[3]Furthermore, as we also separate SZK from PP in communication complexity [64], this implies placing SZK in PP would require non-algebrizing techniques as well [21].

Inspired by this debate, in Chapter 9 we study the computational power of modified theories of quantum mechanics which have been proposed to resolve the paradox. This chapter is based on joint work with Ning Bao and Stephen Jordan, which has been published in Physical Review Letters [46].

Prior work had shown that these modifications all give rise to two negative features – they allow for both superluminal signaling using entanglement, and for the solution of NP-hard problems in polynomial time by speeding up quantum search. The former consequence is considered undesirable by physicists, as it leads to a break down in causality, while the latter is considered unreasonable by computer scientists. We show that these two consequences go hand in hand – in a quantitative fashion, the "amount" of each is polynomially related. Therefore, either consequence is an equivalently valid reason for rejecting these theories. We also show that a modified version of CQP in which one can clone quantum states (which is more powerful than the ability to make non-collapsing measurements) is very powerful, in that it can solve NP-complete problems in polynomial time. The exact power of CQP remains open.

Chapter 2

Preliminaries: The spaces below and above BQP

In this chapter, we introduce the preliminary background required to study the space around BQP. We first define BQP, examine its strengths and weaknesses, and then introduce the notions of quantum advantage from sampling, quantum computing with restricted gate sets, and the power of modified theories of quantum mechanics. We also provide a primer on representation theory and Lie algebras as these will be used on our proofs.

Those already familiar with quantum computing and computational complexity theory can skip to Sections 2.1.4 and 2.2, where we introduce notions specific to gate set universality, quantum advantage from sampling, and restricted gate set models of computation.

2.1 Classical Complexity, Quantum Complexity and BQP

2.1.1 Classical Complexity Bootcamp

To understand the power of quantum computing, we will be using the language of computational complexity theory. The study of computational complexity focuses on asymptotics - i.e. how the amount of computational resources scale with the input size for large inputs. This is denoted by big-O notation - for instance an $O(n)$ problem is one which can be solved in time which scales linearly in the input size. The big-O notation hides constants and subleading terms (like an additive $n^{1/2}$) in the above example. For a broad introduction to computational complexity theory and big-O notation, please see [38].

Much of computational complexity revolves around the study of decision languages. Formally a language $\mathcal{L} \subseteq \{0,1\}^*$ is a subset of strings. The computational complexity of $\mathcal{L}$ is measured by how difficult it is to determine if a string $x \in \{0,1\}^n$ belongs in $\mathcal{L}$. This difficulty is measured in terms of how the time to determine if $x \in \{0,1\}^n$ scales as $n \to \infty$.

Languages of similar "complexity" or "difficulty" are grouped together in complexity classes - which are therefore sets of sets of strings. These are denoted by capital sans-serif letters. For instance, the set of languages decidable in polynomial time is denoted P. The set of languages decidable in polynomial time by randomized algorithms (say with high probability over the coin flips of the algorithm) is denoted BPP.

The complexity classes P and BPP in some way represent the power of "efficient classical computation." However, to many people who are new to computational complexity theory, it may seem odd to classify the power of computational models based on subsets of strings.

Given an input $x \in \{0,1\}^n$, computers can do many things which are more complicated that outputting a yes/no answer in response to an input. For instance, many algorithms have multiple bits as output. Furthermore, there may be many correct possible answers (e.g. "output a factor of the input number) for each input. Instead of defining languages over strings, it seems these sorts of models would define arbitrary *relations* $R \subseteq \{0,1\}^* \times \{0,1\}^*$. One could even allow the computer to output a probability distribution over outputs in response to an input - a notion which we will discuss in detail later.

Given these considerations, why the intense focus on formal languages? There are two reasons. First, decision problems are certainly the simplest model of computation one can study. Therefore, it makes sense to intensely focus on the simplest model first, before trying to classify broader forms of computational problems. Second, oftentimes many problems can be reduced to decision problems. For instance, if you are trying to compute a function with multiple output bits, then you can turn that into several decision problems of the form "output the ith bit of the answer." If one has an efficient algorithm for computing the function, then one can certainly efficiently compute these decision problems as well. Third, the truth is that complexity theorists actually *are* interested in the more broad notions of computation, like relation problems and sampling problems. The reason decision problems take all the spotlight is that they need a simple "poster child" for each model of computation. They can then define broader notions of computation based off each model. For instance, given a complexity class $\mathcal{C}$ of decision languages, the class $\mathsf{F}\mathcal{C}$ denotes those functions (with multiple output bits) computable in the $\mathcal{C}$ model of computation. $\mathsf{Samp}\mathcal{C}$ denotes the probability distributions sampleable in the $\mathcal{C}$ model of computation. So in essence, the decision languages are simply convenient labels for different models of computation.

There are many, many complexity classes that one can define, leading to an "alphabet soup" of computational models [18]. Some of them capture models of "efficient" computation, which one might expect to be instantiated in the real world. For instance, polynomial-time algorithms are captured by the complexity class P. Other classes capture completely unrealistic models of computation. For instance, EXP denotes the set of problems decidable with exponential-time algorithms. Most such algorithms are entirely infeasible to run in the real world, even on small instance sizes. Why study such abstract classes? Surprisingly, there is more reason to do so than mathematical curiosity. It turns out that by studying *unrealistic* models of computation and their relationships between one another, one can learn new facts about *realistic* computational models. In particular, to show that quantum computers have an advantage over classical computation for sampling problems, our proof will go through several non-realistic complexity classes as intermediate steps. This should not be surprising; the phenomena that sometimes abstracting a problem can provide new insights is pervasive in mathematics.

We will now introduce some of the basic classical complexity classes we'll be using in the rest of this thesis. This is an abbreviated crash course on classical complexity for quantum complexity theorists. You may want to pour yourself a coffee before diving in.

P is the set of languages decidable by polynomial time deterministic algorithms. More formally, it's defined as the set of languages $\mathcal{L}$ for which there exists a poly time Turing machine M, which taking input x, accepts if $x \in \mathcal{L}$ and rejects if $x \notin \mathcal{L}$.

BPP is the set of languages decidable by polynomial time randomized algorithms. More formally, it's defined as the set of languages $\mathcal{L}$ for which there exists a randomized poly time Turing machine M, which accepts w.p. $\geq 2/3$ if $x \in \mathcal{L}$, and accepts w.p. $\leq 1/3$ if $x \notin \mathcal{L}$. Here probabilities are taken over the internal coin flips of the randomized algorithm. The choices of the constants $1/3$ and $2/3$ here are arbitrary; so long as their gap is at least inverse

polynomial, by running the algorithm polynomially many times one can determine which is the case with overwhelmingly high probability $1 - 1/\mathsf{poly}$. This "amplification" property is something that will be shared with BQP, which is also a randomized algorithm class since quantum mechanics inherently produces randomness.

P and BPP represent what is efficiently computable in polynomial time in the "real world" by classical computers. Although clearly $\mathsf{P} \subseteq \mathsf{BPP}$ by definition, it is unclear a priori if $\mathsf{BPP} \subseteq \mathsf{P}$, i.e. if every randomized algorithm can be made deterministic. This is widely believed among computational complexity theorists because of the creation of pseudo-random number generators, which assuming certain circuit lower bounds hold, imply that one can create "fake randomness" which is as good a true randomness in the context of BPP algorithms. We refer the interested reader to [200, 156] for details. Therefore in this thesis we will use the term "efficient classical algorithm" or "efficient classical computation" to mean P or BPP algorithms interchangeably.

We will also use a number of "nonrealistic" models of computation in this thesis as well. It turns out that reasoning about such models of computation can be useful in studying the power of "down to earth" realistic quantum computational devices.

NP is the set of languages for which the answer can be verified in polynomial time. A canonical problem in NP is $3SAT$ - i.e. given a set of clauses over n variables, each of which is the OR of three literals (for example $x_1 \vee \bar{x}_2 \vee x_3$, determine if there is a boolean assignment to the variables $x_1 \ldots x_n$ which renders all clauses true. Clearly given such an assignment, one can verify its validity efficiently - simply by checking each clause is satisfied. But finding such an assignment - or determining if one exists - may be very difficult as there are exponentially many such assignments. Although many believe that $\mathsf{P} \neq \mathsf{NP}$ - i.e. some problems in NP such as 3SAT require super-polynomial time to solve - this remains a famously difficult open mathematical problem, and carries a 1 million dollar prize from the Clay foundation for its solution. For a survey of this area see [10].

One strange property of the class NP is that, unlike P and BPP it is not known to be closed under complement. In other words, even if there is an efficient proof that a 3SAT instance is satisfiable (by providing an assignment), there might not be a proof that an instance is unsatisfiable, since this can't be certified by any single assignment. As a result, the complement of NP, denoted coNP, is believed to be a different class.

Note that the set of languages $\mathcal{L} \in \mathsf{NP}$ can be formalized as the set of languages for which these exists a poly-time Turing Machine M such that

$$x \in \mathcal{L} \leftrightarrow \exists y \in \{0,1\}^{\mathsf{poly}(n)} M(x,y) = 1$$

In other words, there exists a proof such that the verification algorithm M accepts the proof as valid.

On the other hand, coNP languages consist of those problems (say those 3SAT instances) for which *no* proof exists. In other words, this could be formalized as

$$x \in \mathcal{L} \leftrightarrow \forall y \in \{0,1\}^{\mathsf{poly}(n)} M(x,y) = 0$$

By flipping the output of M at the end of the computation, this can be equivalent to stating that $\mathcal{L} \in \mathsf{coNP}$ if there exists a poly-time verification algorithm M' such that

$$x \in \mathcal{L} \leftrightarrow \forall y \in \{0,1\}^{\mathsf{poly}(n)} M'(x,y) = 1$$

Therefore the definition of coNP looks exactly like the definition of NP, except the exists

quantifier over proofs is replaced by a for all quantifier.

The Polynomial Hierarchy, denoted PH, a generalization of both NP and coNP, which works by adding additional quantifiers on to these. It is defined as follows: The class Σ_1 is defined as NP. The class Π_1 is defined to be coNP. The class Σ_2 is defined as those languages for which there exists a poly-time Turing Machine M such that

$$x \in \mathcal{L} \leftrightarrow \exists x_1 \in \{0,1\}^{\mathsf{poly}(n)} \forall x_2 \in \{0,1\}^{\mathsf{poly}(n)} M(x, x_1, x_2) = 1$$

In other words, it looks exactly like the definition of coNP = Π_1, but with an additional existential quantifier over an additional poly-sized proof. You could summarize this as $\Sigma_2 = \exists.\Pi_1$ Π_2 is defined similarly, except the for all and exists quantifiers are swapped - i.e. it is $\forall.\Sigma_1$. Σ_k and Π_k are defined recursively in this manner; Σ_k having k quantifiers starting with $\exists$, and Π_k starting with $\forall$.

Clearly $\Pi_k \subseteq \Sigma_{k+1}$, and $\Sigma_k \subseteq \Pi_{k+1}$. So these two towers of classes interweave into one another. PH is then defined to be $\cup_{k \in \mathbb{Z}} \Sigma_k \cup \Pi_k$ - i.e. the union of these entire towers of classes.

One can think of levels of PH as generalizations of NP and coNP. Each level's relation to the level below is the same relation as P is to NP or to coNP. Therefore just as complexity theorists believe that P $\neq$ NP, it is often widely conjectured that the Polynomial Hierarchy is infinite - i.e. adding each additional quantifier increases the power of the complexity class. This assumption is known as the "non-collapse" of the polynomial hierarchy[1].

The last remaining classes we will need are counting classes. While NP captures the problem of deciding if a solution to a 3SAT instance exists, these counting classes will capture the complexity of *counting* the number of solutions. More formally, the class #P is defined to be the set of functions $f : \{0,1\}^* \rightarrow \{0,1\}*$ for which there exists a poly-time Turing Machine M such that $|\{y \in \{0,1\}^{\mathsf{poly}(n)} : M(x,y) = 1\}| = f(x)$. A canonical example of such a problem is #SAT - i.e. counting the number of satisfying assignments to a 3SAT instance.

Clearly #P problems are not decision problems - because they have multiple output bits. Therefore one cannot compare #P to decision classes like P or NP. However, one can define a decision analogue of #P, which is the class PP. PP solves the decision problem of deciding if the number of satisfying assignments is above or below a certain threshold t. One can also view PP as the set of languages decidable by randomized algorithms with probability strictly $> 1/2$. In other words, the PP machine accepts if more than half of the possible coin flip outcomes cause the randomized Turing machine to accept, and rejects otherwise. This is easily seen to be equivalent to the prior definition.

In short we've introduced two "non-realistic" models of computation - the polynomial hierarchy on the one hand based on alternating quantifiers, and counting problems (PP) on the other. Which is more powerful? It turns out that, in essence, counting is more powerful than alternating quantifiers. In particular, Toda showed that PH $\subseteq$ P^{PP} [232]. Here the notation A^B means that A has the ability to solve instances of B at unit cost. B is referred to as the "oracle" as it "magically" solves instances of B. So P^{PP} means a classical algorithm with the ability to decide PP problems - i.e. with the ability to count solutions - at unit cost. Therefore, if one has the ability to count as a subroutine, it is more powerful than any level of the polynomial hierarchy. This fact will prove critical the later chapters of this

[1]The reason for this terminology is that if any level is contained in any other - e.g. if $\Sigma_k = \Sigma_{k-1}$ - then in fact all the complexity classes in PH are equal to Σ_{k-1}. Therefore the infinite tower of classes "collapses" to a single class. For instance if P = NP then the entire hierarchy collapses to PH = P.

thesis.

2.1.2 Quantum computing basics and BQP

This section will describe the basics of quantum computing, culminating in the definition of BQP. BQP represents those decision problems which can be solved in polynomial time by quantum computers. But to explain what this means, we need to first define how quantum computing works. For a more detailed explanation, see [199].

The easiest way to understand quantum computing is by analogy to randomized classical computing. We'll first describe what will at first seem like a strange model of classical computation, which is equivalent to BPP. We'll then generalize this to quantum computation.

A probability distribution over n-bit strings can be specified by a vector $v \in (\mathbb{R}^+)^n$ of nonnegative probabilities, which sum to 1, i.e. $|v|_1 = 1$. We can think of the state of a randomized algorithm as always being in such a probability distribution (where the probabilities are taken over the choice of random seed for the algorithm). At the beginning of the algorithm, we can assume the probability distribution is as follows: the state of the first n input bits are the input $x \in \{0,1\}^n$ with certainty, and additionally there are poly(n) input bits in a uniformly random state (these are the coin flips to be used by the randomized algorithm). The net probability distribution on these $n + \text{poly}(n)$ bits is given by the tensor product[2] of the distribution on the first n bits with that on the last poly(n) bits.

Now let's model a randomized computation according to a classical circuit on this (random) input. A classical circuit breaks down a polynomial time computation in polynomially many simple steps - such as individual logical gates like AND and NOT acting on small numbers of bits. So long as the circuit is "uniformly generated" - i.e. produced by a poly-time algorithm - then this model of computation is equivalent to BPP.

Each logic gate can be specified by a matrix whose rows are indexed by inputs to the gate, and whose columns are indexed by outputs. For instance, a NOT gate can be represented by the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. If one takes a probability distribution over a bit and represents it as the vector (p_0, p_1), then applying this matrix to this vector gives the new probability distribution on the bit after the gate is applied. This reduces gate application to matrix multiplication. So long as each row of the matrix has exactly one "1" in it (and the rest 0's) - i.e. so long as the gate has a deterministic output given the input - this maps probability distributions to probability distributions. Note that when one applies a gate to some subset of the bits, the matrix action on the entire probability vector v is given by the tensor product of the gate on the relevant bits with the Identity on the remaining bits[3]. This performs the intuitive action of "leaving the other bits alone" while applying the gate to a subset of gates only.

At the end of the computation, we say the circuit "accepts" if the probability the first output bit is "1" is more than 2/3. This probability is given by summing over exactly half of the probabilities in v. In this fashion, BPP can be defined as the set of problems decidable by uniformly generated randomized circuits. Less trivially, one can also assume without loss of generality that the circuit is reversible [233] - in other words each gate consists of

[2]Recall that operationally, the tensor product of two vectors is obtained by multiplying all possible combinations of entries in each vector in order. For instance $(a,b) \times (c,d)$ is given by (ac, ad, bc, bd). Here we are assuming that the probabilities are listed in the binary order of the underlying strings.

[3]Operationally, the tensor product of two matrices $A \otimes B$ is obtained in a similar fashion to the tensor product for vectors. For example $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \otimes \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae & af & be & bf \\ ag & ah & bg & bh \\ ce & cf & de & df \\ cg & ch & dg & dh \end{pmatrix}$.

a permutation of inputs and outputs. Although AND and OR gates are not reversible, so can't be used in reversible circuits, it turns out that the Toffoli gate - i.e. the controlled-controlled-NOT gate, which flips the third bit iff the first two bits at 1, is universal for classical computing in this manner.

Now that we have given a rather convoluted definition of classical randomized computation, we can define what we mean by quantum computation. The definition is essentially the same, with several minor changes. First, the state of n "qubits" is specified by a vector $v \in (\mathbb{C}^2)^{\otimes n}$ such that $||v||_2 = 1$. In other words, each possible string is assigned a possibly complex "amplitude", and these amplitudes are normalized in the ℓ_2 norm (rather than the ℓ_1 norm as in classical probability). Initially the state is in the state $v_{0^n} = 1$ and otherwise $v_x = 0$. This state is denoted $|0^n\rangle$; in general the state $|x\rangle$ denotes the unit vector $v_x = 1$ and all other entries 0. So arbitrary quantum states take the form $|\psi\rangle = \sum_x \alpha_x |x\rangle$ where $\sum_x |\alpha_x|^2 = 1$ The basic operations allowed are *unitary* matrices acting on a small number of qubits at a time, in tensor product with the identity on the remaining qubits. A unitary is a matrix M such that $MM^\dagger = I$ - with possibly complex entries. Such operations manifestly preserve the ℓ_2 norm of the vector. For now assume that one can apply any two-qubit unitary. It might be disconcerting to some that there are a continuum of unitaries, but we will see later that without loss of generality one can use discrete sets of quantum gates as well. After polynomially many gates, the state is "measured", and outcome x is sampled with probability $|v_x|^2$. We say the circuit accepts if the probability the first qubit is 1 is more than $2/3$, and that it rejects if the probability the first qubit is 1 is less than $1/3$.

In short, BQP is defined identically to BPP, except that the ℓ_1 norm is replace by the ℓ_2 norm, and the gates are those matrices which preserve the ℓ_2 norm, rather than gates which preserve the ℓ_1 norm. A more formal definition is below.

Definition 2.1.1. A language $\mathcal{L} \in$ BQP iff there exists a classical poly-time Turing Machine M such that for any $x \in \{0,1\}^n$, $M(x)$ outputs a classical description of a quantum circuit C consisting of polynomially many two-qubit gates, such that if $|\psi\rangle$ is the state obtained by applying C to the state $|0^n\rangle$,

- If $x \in \mathcal{L}$, then upon measurement the first qubit of ψ is 1 w.p. $\geq 2/3$
- If $x \notin \mathcal{L}$, then upon measurement the first qubit of ψ is 1 w.p. $\leq 1/3$

We have defined BQP by using circuits on qubits - i.e. circuits over the Hilbert space $(\mathbb{C}^2)^{\otimes n}$. However note that there are many equivalent definitions of BQP defined over other Hilbert space, for example topological quantum computing [120], or adiabatic quantum computing [110, 32].

2.1.3 Strengths and weaknesses of BQP

Now that we have defined BQP, it is natural to study how BQP compares to other classical complexity classes. We will see that BQP is an "odd duck" in the complexity landscape.

First, what is the evidence that BQP is bigger than BPP? We know that the problem of factoring integers lies in BQP - this is a consequence of Shor's algorithm [221]. We also know that BQP contains several other languages not known to be in BPP - such as simulating quantum mechanics, and approximating the Jones polynomial of a knot at certain roots of unity [31]. These can be seen as evidence that BQP is larger than BQP. However, thus far

there are no known negative complexity-theoretic consequences of either of these problems lying in BPP.

It is widely believed that BQP does not contain NP. This belief is based on two facts. First, there is an oracle relative to which BQP does not contain NP. This follows from the lower bound for Grover search [56], which states that searching a black-box list of N items requires $\Omega(N^{1/2})$ queries to the black box. Therefore a quantum computer cannot "brute-force" search over the list of possible assignments to a 3SAT instance in order to solve it, while an NP machine can do so by definition. Second, there is thus far no quantum algorithm for an NP-complete problem, so it is conjectured that even quantum computers require exponential time to solve 3SAT. In the other direction, Watrous [239] gave an oracle relative to which BQP is not contained in NP (or even MA). In general it seems difficult to verify the results of generic quantum computations. The known BQP-complete problems are not known to lie in NP. Therefore BQP may not lie in NP either. In fact it is conjectured that there is an oracle relative to which BQP is not contained in PH [7, 113], though thus far this problem remains open, although the relational versions of the classes have been separated [7].

Likewise, there are oracle separations between BQP and other weak models of computation - for instance there are oracles relative to which BQP is not in SZK [7, 89] and vice versa [2].

The best known classical upper bound for BQP is AWPP which is contained in PP [119][24]- i.e. those decision languages which represent the power of counting problems. This is a very powerful complexity class - recall that P^{PP} contains all of PH [232]. It remains open to improve this upper bound.

In short BQP seems to be "just a bit bigger" than BPP, but in a somewhat "new direction" in hardness - in the direction of PP but incomparable to NP or SZK. The relationship between BQP and these complexity classes is given in Figure 2-1.

2.1.4 Gate sets and universality

In our definition of BQP in Section 2.1.2, we allowed the quantum circuit to have gates drawn from the continuous set of two-qubit gates. In this section, we show how to relax this definition to allow for a discrete set of gates, using the Solovay-Kitaev Theorem.

We say a discrete set of quantum gates $G = \{g_1 ... g_k\}$, each of which act on k qubits, is *universal on k' qubits* if they densely generate all unitaries on k' qubits. In other words, for every unitary $U \in U\left((\mathbb{C}^2)^{\otimes k'}\right)$, and any $\varepsilon > 0$, there exists a finite sequence of gates $g_{i_1} g_{i_2} \dots g_{i_l}$ which ε-approximate U in the operator norm. For shorthand we will often simply write G is universal to denote that there exists a d' for which is it universal on d' qubits. First a few comments on this definition. First, note that this definition composes. Say that a gate set G is universal, and a gate set G' densely generates G. Then clearly G' must be universal as well. For this reason, as $k' \geq 2$, then universality on k' qubits implies universality on $k'' > k'$ qubits as well. This is because it is known that the set of all two qubit gates is universal on arbitrary numbers of qubits [199], so by composing these statements one obtains the result. Second, note that this definition says nothing about the efficiency of generating all unitaries - i.e. the lengths of the sequences of gates required to achieve accuracy ε.

The Solovay-Kitaev theorem [99] is a general technique for turning statements about density into statements about efficiency. Specifically it says the following:

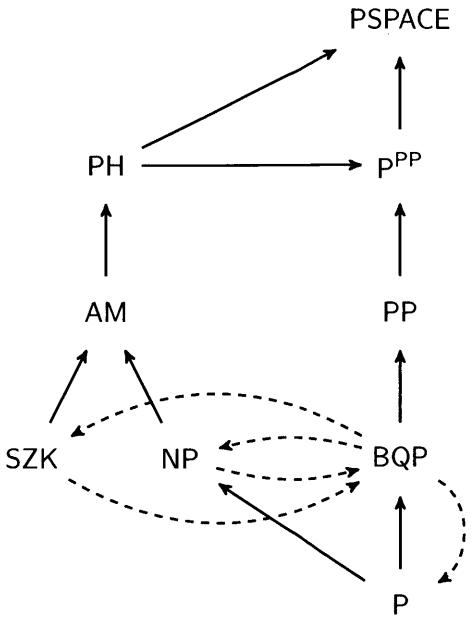


Figure 2-1: Known relationships between BQP and nearby complexity classes. $\mathcal{C}_1 \to \mathcal{C}_2$ indicates $\mathcal{C}_1$ is contained in $\mathcal{C}_2$ respect to *every* oracle, and $\mathcal{C}_1 \dashrightarrow \mathcal{C}_2$ denotes that there is an oracle $\mathcal{O}$ such that $\mathcal{C}_1^{\mathcal{O}} \not\subset \mathcal{C}_2^{\mathcal{O}}$. Only oracle separations involving BQP are drawn on the diagram; other separations (such as AM not in PP relative to an oracle [238]) are omitted.

Theorem 2.1.2 (Solovay-Kitaev Theorem [99]). *If a set of gates G is universal on k qubits, then for any $\varepsilon > 0$, one can ε-approximate arbitrary unitaries on k qubits using a sequence of gates from G and their inverses of length*

$$O(2^k \log^{3.97}(1/\varepsilon))$$

Furthermore there is an efficient algorithm to generate such a sequence.

This remarkable result shows that one can efficiently translate between different universal gate sets. If one wishes to change gate sets in the definition of BQP, by standard amplification it suffices to achieve constant error in the compilation of the circuit, which means inverse polynomial error on each individual gate. The Solovay-Kitaev Theorem therefore says that one can change between universal gate sets for BQP while merely losing polylogarithmic factors. Hence the choice of universal gate set for BQP is irrelevant from the perspective of asymptotic complexity.

There have been many improvements and generalizations of the Solovay-Kitaev Theorem. For instance, several works have improved the exponent on the logarithm to 1 for particular gate sets of experimental interest, and furthermore decreased the leading constant to a very small number. For example, see [209, 60] and references therein. The theorem has also been improved to cover non-unitary matrices as well [29]. However one improvement which has so far eluded researchers is removing the assumption that one can apply exact inverses of the gates as well. Sardharwalla et al. [212] recently showed that one can get away without inverses, so long as the gate set contains the Weyl group (which on qubits is the Pauli group). But proving a general inverse-free Solovay-Kitaev theorem remains open. This will

significantly complicate our results in Chapters 4 and 5.

While it has some limitations, the Solovay-Kitaev theorem allows you to change between universal gate sets efficiently. There are many universal gate sets known. For instance, Controlled-NOT plus the set of one qubit gates is universal [49]. Controlled-NOT, Hadamard, and a $pi/8$ phase gate are universal as well [169]. The Toffoli gate (Controlled-Controlled-NOT) plus any basis-changing real gate are universal [220]. In fact, Lloyd [180] and others [102, 240, 90, 52] have shown that a Haar-random two-qubit gate is universal with probability 1. So "most" gate sets are universal. Overall it seems universality is difficult to avoid. The Solovay-Kitaev Theorem implies that universal gate sets are all computationally equivalent.

Note that by a result of Ivanyos, given a set of gates G with algebraic entries, it is decidable whether or not the gate set is universal [157]. Impressively, as part of the proof Ivanyos shows that any gate set of d-qubit gates which is not universal on $d' \approx 2^8 d$ qubits will never become universal for any larger d'. Therefore the problem of deciding universality reduces to merely checking if all operations are generated on a large but finite number of qubits. A different algorithm which decides universality was recently given by Sawicki and Karnas [214], which is guaranteed to terminate in a finite number of steps for any gate set, but without an upper bound on the number of steps needed.

Despite the fact that we know most gate sets are universal, the problem of proving universality for *particular* gate sets - or proving classification theorems showing any gate set of a particular form is universal - is quite hard[4]. The primary difficulty in proving universality is in proving that the gate set generated is infinite. Once that has been achieved, techniques from Lie algebras can be applied to the problem to complete the proof. It suffices to show the Lie algebra contains all of $\mathfrak{su}(2^d)$, or one can apply the more compact criteria of Karnas and Sawicki [214]. There are many examples of such uses of Lie algebraic techniques to prove universality once the group is promised to be infinite, for example [202, 214, 66, 213].

There are several sufficient criteria which are used to prove the gate set generated is infinite. The first is to find a rotation by an irrational multiple of π. Often this is achieved using algebraic number theory. In particular twice the cosine of a rational multiple of π is an algebraic integer (this follows from the multiple angle formula for cosines), and hence one can appeal to special properties of algebraic integers to prove (by contradiction) that an angle is irrational. For instance this technique was used by Yaoyun Shi to show that Toffoli plus any real non-basis-preserving gate is universal[5] [220]. The second technique is to enumerate all possible discrete subgroups of the parent group, and show the gate set cannot generate a representation of any of them, as we will do in Chapter 3 [63]. This is a very tedious task however, and can only be used for small-dimensional unitary groups for which the finite subgroups have been classified, i.e. only for dimension 3 or 4 [109, 139, 142]. Third, Karnas and Sawicki recently gave a sufficient criterion for a group to be infinite [214] and used this to reproduce the classification of finite subgroups of $SU(2)$. Fourth, one can appeal to the result that the Clifford group plus any non-Clifford element is universal; this result is based on complicated arguments involving invariants of the Clifford group [196, 197].

[4]This is somewhat analogous to the fact that proving lower bounds for the circuit complexity of particular functions is hard, despite the fact that we know most functions have high circuit complexity.

[5]In particular Shi used the result of Włodarski [243] as a black box, but Włodarski's result uses algebraic number theory.

2.2 Proving quantum advantage from sampling problems

In this section, we define what it means to have a quantum advantage for a sampling problem. We show how one can generically obtain hardness of (exactly) sampling from the output distribution of (noiseless) quantum computers, and discuss more realistic noise models.

2.2.1 Why sampling?

One reason that quantum complexity theorists have turned their attention to sampling problems, rather than decision problems, is that it is very difficult to establish quantum advantage for decision problems. First, proving that BQP is bigger than P would show that P $\neq$ PSPACE, which is well beyond our current reach. A more modest goal is to establish quantum advantage under widely accepted complexity assumptions like P $\neq$ NP, non-collapse of the polynomial hierarchy PH, and others. However, even this problem has remained open. The difficulty is that the decision languages known to be in BQP, but not known to be in P or BPP are not known to be NP-hard[6]. So placing these problems in P is not known to imply P = NP or collapse PH, or other similar negative consequences in complexity. Therefore, thus far no one knows how to derive BQP $\neq$ BPP from assumptions such as P $\neq$ NP or PH being infinite.

Furthermore, there are practical difficulties with proving quantum advantage from decision problems such as factoring. First, the quantum resources needed to factor integers are experimentally challenging to implement. For instance, Shor's algorithm requires long-range interactions between qubits, while most experimental devices only couple nearest-neighbor qubits on a certain geometry. It also requires a universal set of quantum gates[7]. Furthermore the number of qubits needed to factor an n-bit integer is currently $2n + 1$ [123, 144]. To date the largest number factored by Shor's algorithm[8] is 21 [187, 237]. This is much smaller than the largest numbers factored by classical computers to date, as nontrivial factoring algorithms which are in the hundreds of digits due to nontrivial factoring algorithms [84]. Factoring such numbers is beyond the capability of near-term quantum devices, which have in the tens of qubits [193] Therefore if one is hoping to provide an empirical demonstration of quantum advantage (sometimes referred to as "quantum supremacy") - i.e. perform a quantum experiment which could not be simulated by any classical computer to date [207, 61, 15] - then we are very far from achieving this with factoring.

More recently, several "weak" models of quantum computation have been proposed which appear to achieve hardness at smaller numbers of qubits or particles (usually around 50). These constructions have worked by considering a broader notion of computation known as sampling problems, where the goal is to simulate the entire output of the quantum device, rather than simply the output of the first qubit. In these constructions, it is shown that these devices can sample from some *probability distributions* which classical computers cannot efficiently sample from exactly, assuming PH is infinite, and also can't be sampled to reasonable error under additional assumptions. That is, there exists a family of probability distributions D_x labeled by strings $x \in \{0,1\}^*$, such that given input x quantum devices can sample from D_x efficiently, but randomized classical computers cannot. Fur-

[6]For instance, Factoring lies in NP $\cap$ coNP, so it cannot be NP-hard unless PH collapses.

[7]On the bright side however, Shor's algorithm can be made to have only logarithmic circuit depth [92], which is helpful to experimentalists.

[8]Note that these experiments used "compiled" versions of Shor's algorithms which presumed knowledge of the factors, and hence might not be considered full demonstrations of Shor's algorithm [223].

thermore, these models have the advantage that they require fewer experimental resources than universal quantum computation. Therefore such experiments could help provide the first demonstration of quantum advantage.

2.2.2 Notions of simulation

In order to define sampling tasks, we will need to define what it means to "simulate" a quantum device. There are several different notions one can consider.

The results of this thesis primarily focus on the notion of (approximate) *weak* simulation. A *weak* simulation of a family of quantum circuits is a classical randomized algorithm that samples from the same distribution as the output distribution of the circuit. On the other hand, a *strong* simulation of a family of quantum circuits is a classical algorithm that computes not only the joint probabilities, but also any marginal probabilities of the outcomes of the measurements in the circuit. Following [173], we can further refine these definitions according to the number of qubits being measured: a strong(1) simulation computes the marginal output probabilities on individual qubits, and a strong(n) simulation computes the probability of one particular output string $y \in \{0,1\}^n$. Likewise a weak(1) simulation simulates the output on a single qubit, and a weak(n) simulation simulates the output on all n output qubits. Unless otherwise specified, here "weak" and "strong" simulation will refer to weak(n) and strong(n) simulations, respectively - i.e. we are considering the complexity of simulating all outputs.

Furthermore, within the space of weak(n) simulations of quantum devices, we will make frequent use of the following notions of *approximate* weak simulations. More formally, let $\mathcal{P} = \{p_z\}_z$ and $\mathcal{Q} = \{q_z\}_z$ be (discrete) probability distributions, and let $\varepsilon \geq 0$. We say that $\mathcal{Q}$ is a *multiplicative ε-approximation* of $\mathcal{P}$ if for all z,

$$|p_z - q_z| \leq \varepsilon p_z. \tag{2.1}$$

We say that $\mathcal{Q}$ is an *additive ε-approximation* of $\mathcal{P}$ if

$$\frac{1}{2}\sum_z |p_z - q_z| \leq \varepsilon. \tag{2.2}$$

Note that any multiplicative ε-approximation is also an additive $\varepsilon/2$-approximation, since summing Eq. (2.1) over all z produces Eq. (2.2). Here the factor of $1/2$ is present so that ε is the total variation distance between the probability distributions.

A *weak simulation with multiplicative (additive) error* $\varepsilon > 0$ of a family of quantum circuits is a classical randomized algorithm that samples from a distribution that is a multiplicative (additive) ε-approximation of the output distribution of the circuit. Note that from an experimental perspective, additive error is the more appropriate choice, since the fault-tolerance theorem merely guarantees additive closeness between the ideal and realized output distributions [30].

2.2.3 Quantum advantage from sampling problems

In this section, we will show how to prove hardness of sampling results for restricted models of quantum circuits assuming non-collapse of PH. Specifically, suppose that one considers a "weak" model of quantum computing $\mathcal{C}$, defined as poly-sized circuits using some *non-universal* gate set G. Let $\mathsf{Samp}\mathcal{C}$ denote the set of sampling problems efficiently sampleable

by $\mathcal{C}$. We will show how to prove that the class $\mathcal{C}$ cannot be weakly simulated (i.e. weak(n) simulated) in SampBPP to constant multiplicative error unless PH collapses. Therefore no classical device can simulate the probability distributions output by such circuits (assuming non-colapse of PH). As we will see later in the thesis, this can occur even if the decision problems in $\mathcal{C}$ lie in P - i.e. even if such circuits can be strong(1) simulated.

The below arguments were first given by Aaronson and Arkhipov [11] and Bremner, Jozsa & Shepherd [72], but we recap them here as they will be used several times in this thesis.

In order to reason about the computational complexity of Samp$\mathcal{C}$ distributions, we will need to introduce the idea of postselected circuits, which we will relate to classical complexity classes such as PP.

A postselected quantum circuit is a circuit where one specifies the value of some measurement results ahead of time, and discards all runs of the experiment which do not obtain those measurement outcomes. This is not something one can realistically do in a laboratory setting, because the measurement outcomes you specify may occur extremely infrequently—in fact, they may be exponentially unlikely. However, postselection can help you examine the *conditional probabilities* found in the output distribution of your circuit. In particular, if you can show that those conditional probabilities can encode the answers to very difficult computational problems, then this can provide evidence against the ability to classically simulate such circuits. Therefore, we will now define what it means for a set of probability distributions to decide a problem under postselection. The basic idea is that if some of the conditional probabilities of the system encode the answer to a problem, then we say that problem can be decided by postselected versions of these probability distributions. We define this more formally below:

Definition 2.2.1. Let Post$\mathcal{C}$ be the set of languages $L \subseteq \{0,1\}^*$ for which there exists a family of $\mathcal{C}$ circuits $\{\mathcal{D}_x\}$ and a classical poly-time algorithm which, given an input length n, outputs a subset B of qubits and a string $z \in \{0,1\}^{|B|}$ such that

- If $x \in L$, then Pr[$\mathcal{D}_x$ outputs 1 on its first bit | bits B take value z] $\geq 3/4$.
- If $x \notin L$, then Pr[$\mathcal{D}_x$ outputs 1 on its first bit | bits B take value z] $\leq 1/4$.

In other words, there exists a poly-time algorithm which outputs an experimental setup and a postselection scheme such that the conditional probabilities of $\mathcal{D}_x$ encode the answer to the problem. In general, the choice of constants $1/4$ and $3/4$ in the above definition might matter. For instance, when Post$\mathcal{C}$ is not capable of universal classical computation, it is unclear how to take the majority vote of many repetitions to amplify the success probability. However, we only consider the class Post$\mathcal{C}$ in cases where Post$\mathcal{C}$ can perform universal classical computation, and thus the choice of constants $1/4$ and $3/4$ is arbitrary.

One can likewise define the classes PostBQP and PostBPP[9] which capture the power of postselected quantum computation and postselected randomized computation, respectively.

One of the main technical tools we will use in our proof is the following lemma, which was first shown by Bremner, Jozsa, and Shepherd [72], but which we will make extensive use of in our thesis:

Lemma 2.2.2. *Suppose that* PostBQP $\subseteq$ Post$\mathcal{C}$ *for some restricted model of quantum computation* $\mathcal{C}$. *Then* BPP *machines cannot weakly simulate* Samp$\mathcal{C}$ *with multiplicative error* $c < 1/2$ *unless* PH *collapses to the third level.*

[9] PostBPP is more commonly known as $\mathsf{BPP}_{\mathsf{path}}$.

In other words, if postselected $\mathcal{C}$ circuits are capable of performing (postselected) universal quantum computation, then they cannot be weakly simulated by a classical computer to constant multiplicative error under plausible complexity assumptions. The fundamental reason this is true is that the class PostBQP is substantially more powerful than the class PostBPP. In particular, Aaronson [6] showed that PostBQP = PP. Recall from Section 2.1.1 that PP (which stands for Probabilistic Polynomial-time) is the set of languages L decidable by a poly-time randomized algorithm M, such that

- If $x \in L$, then $\Pr[M(x) \text{ accepts}] > 1/2$;
- otherwise, $\Pr[M(x) \text{ accepts}] \leq 1/2$.

In other words, the class PP represents the class of problems solvable by randomized algorithms, where the probability of acceptance of "yes" and "no" instances is different, but may only differ by an exponentially small amount[10]. Toda's Theorem [232] states that $\mathsf{PH} \subseteq \mathsf{P}^{\mathsf{PP}}$. In other words, the class PP is nearly as powerful as the entire polynomial hierarchy.

On the other hand, the class PostBPP is far weaker; it lies in the third level of the polynomial hierarchy by Stockmeyer's approximate counting theorem[11] [225]. So if one assumes that PH does not collapse to the third level, then PostBPP $\neq$ PostBQP; i.e. PostBQP is a stronger complexity class than PostBPP.

From this, we can now state why the inclusion PostBQP $\subseteq$ Post$\mathcal{C}$ implies there cannot exist an algorithm to simulate Post$\mathcal{C}$ circuits (assuming the non-collapse of PH). Suppose there were a BPP algorithm to weakly simulate such circuits. Then, by postselecting this BPP algorithm, we could solve a PostBQP-hard problem in PostBPP, which would imply the collapse of the polynomial hierarchy. A more formal statement of this proof is given below:

Proof of Lemma 2.2.2. The proof of this corollary is given in [72] Theorem 2 and Corollary 1, but we provide a summary for completeness. Suppose that a BPP machine M can weakly simulate Samp$\mathcal{C}$ circuits to multiplicative error $c < 1/2$. Then for any individual output string x, we have $\frac{1}{2}\Pr[M \text{ outputs } x] < P(x) < \frac{3}{2}\Pr[M \text{ outputs } x]$. Since PostBQP $\subseteq$ Post$\mathcal{C}$, and PostBQP = PP [6], this can be shown to imply PP $\subseteq$ PostBPP. But PostBPP $\subseteq$ PostBQP = PP, so this implies PostBPP = PP. Hence by Toda's theorem [232], we have $\mathsf{PH} \subseteq \mathsf{P}^{\mathsf{PP}} = \mathsf{P}^{\mathsf{PostBPP}} \subseteq \Delta_3$, where Δ_3 is the third level of the polynomial hierarchy. Hence $\mathsf{PH} = \Delta_3$ as claimed. □

Note that in certain cases, Fujii et al. [122] showed that this hardness of simulation result could be improved to imply the collapse of PH to the second level rather than the third, using a different complexity-theoretic argument involving the class NQP. However, their argument is gate-set dependent, so does not apply to generic restricted gate sets.

Therefore, in order to show sampling hardness to constant multiplicative error for Samp$\mathcal{C}$, all one needs to do is to show that Post$\mathcal{C}$ contains PostBQP = PP. In order to do this, a first step is to show that BQP $\subseteq$ Post$\mathcal{C}$ - i.e. that postselected $\mathcal{C}$ circuit are capable of universal quantum computation. This is often achieved by creating various postselection gadgets which perform quantum operations which boost $\mathcal{C}$'s gate sets to universality. Then by postselecting the circuit further, one can show that PostBQP $\subseteq$ Post$\mathcal{C}$. One subtlety with this approach is that the postselection used by Aaronson to show PP = PostBQP only

[10] Note the difference is probabilities is always at least $2^{-\text{poly}(n)}$, because a PP algorithm can only make polynomially many coin flips.

[11] In fact even $\mathsf{P}^{\mathsf{PostBPP}}$ lies in the third level of PH.

succeeds with inverse exponential probability. Therefore, to show PostBQP $\subseteq$ Post$\mathcal{C}$, one must not only show that BQP is contained in Post$\mathcal{C}$, but furthermore one must be able to weak(n) simulate all of the outputs of the BQP circuit to inverse exponential additive error. This is achievable, for instance, if the gadgets used obey a Solovay-Kitaev theorem. From this one can show the following Lemma:

Lemma 2.2.3. *Consider a restricted model of quantum computation C defined be a restricted gate set. Then if one can create postselection gadgets H out of gates from C such that conditioned on the postselection succeeding:*

- *The gates of H combined with the gates of C are universal.*
- *The gates of H combined with the gates of C obey a Solovay-Kitaev theorem.*

Then PostBQP $\subseteq$ Post$\mathcal{C}$ *so by Lemma 2.2.2* Samp$\mathcal{C}$ *cannot be weakly efficiently classically simulated to constant multiplicative error unless* PH *collapses.*

We will use this technique several times in this thesis, in Chapters 4 and 5. However, using this Lemma requires that one can perform the inverses of the gadgets exactly (or possibly with inverse exponential error), since the generic Solovay-Kitaev Theorem requires the gates to be closed under inversion. This can be a major obstacle to completing such proofs, as there is usually not an obvious way to invert postselection gadgets (even if C's gate set is closed under inversion) since postselection is an irreversible operation. One can also get around this by using the inverse-free Solovay-Kitaev Theorem [212] in the special case where the Pauli group belongs to the gadget set - we will use this trick in Chapter 5. However, for this reason we believe that proving an inverse-free Solovay-Kitaev Theorem is a worthwhile open question, since it may make it easier to classify the power of intermediate gate sets.

In the presence of noise, error-corrected quantum computers are capable of sampling from distributions within constant additive error of the ideal quantum circuit with polylogarithmic overhead; this is the fault-tolerance theorem [30]. Therefore proving hardness to constant additive error is a more physically relevant model of error. Most hardness of sampling results for constant multiplicative error can be extended to hardness results for inverse exponential additive error[12]. This can theoretically be achieved by error-correction with polynomial multiplicative overhead, but error-correction is difficult to implement on experimental devices due to the large overheads involved. In fact, near-term quantum supremacy experiments are aiming to perform sampling problems without any error-correction [193], and thus will be subject to constant additive error. Therefore, proving hardness of sampling results to constant additive error is more experimentally relevant in the near term. However, proving hardness of sampling to constant additive error is a more difficult task than multiplicative error. We will see an example of such a proof in detail in Chapter 5, but we outline how these proofs work here. These proofs usually make use of three ingredients. First, they prove an anticoncentration theorem, which states that over the choice of output string $y \in \{0,1\}^n$ and the circuit choice of $\mathcal{C}$, most of the output probabilities are at least a certain amount. Second, they use Markov's inequality to argue that the simulation error is small on most of the output probabilities as well. On the overlap between these sets S,

[12]This is because most reasonable quantum gate sets have the property that the smallest non-zero probabilities in their output distributions are at least inverse exponential [176, 9]. Hence a simulation $\mathcal{S}$ with inverse exponential additive error would still yield a multiplicative gap in the accept/reject probabilities if one postselects $\mathcal{S}$, and the hardness argument carries through.

one therefore has small error and a large output probability - implying that the additive simulation algorithm is also a multiplicative simulation on these outputs. Hardness is then obtained by assuming that most of the outputs are #P-hard to compute to multiplicative error - so in particular some elements of S are hard to compute in this way. From this they obtain hardness of sampling to additive error unless PH collapses. Thus far these average-case hardness assumptions have remained unproven, and it is a major open question to prove one of them. Such a result would require non-relativizing techniques, due to an oracle obstruction by Aaronson and Chen [15]. But such a result would imply that even near-term noisy quantum circuits are hard to simulate unless PH collapses. This would be the strongest evidence yet for quantum advantage over classical computation.

2.2.4 Weak models of quantum computing and hardness of sampling results

A number of authors have proven hardness of sampling results for various forms of "weak" quantum computing. We summarize these results here.

We first list those results which make use of "weak" fragments of quantum computing, which are not known are believed to be capable of universal quantum computing.

Constant depth circuits: In 2004 Terhal and DiVencenzo showed that constant-depth quantum circuits cannot be efficiently classically simulated unless $\mathsf{BQP} \subseteq \mathsf{AM}$ [230]. This would be a surprising conclusion as it is believed BQP may lie outside all of PH [113, 7], while AM lies in the second level of PH [39].

Boson Sampling: In 2010 Aaronson and Arkhipov [11] showed that certain experiments in linear optics are capable of performing hard sampling tasks. In particular, they proved that a linear optical experiment with n identical photons and $m = \mathsf{poly}(n)$ modes cannot be weakly simulated to constant multiplicative error unless PH collapses. They furthermore showed that under two additional conjectures - an anticoncentration conjecture, and a hardness-on-average conjecture - that such optical systems are hard to simulate to constant *additive* error as well. We will show in Chapter 3 that restricting the optical gate sets available to the model leaves the power unchanged [63].

Commuting Circuits (IQP): In this model, one considers the power of quantum computing with a commuting gate set, which is manifestly non-universal. In 2010 Bremner, Jozsa and Shepherd [72] showed that circuits consisting of gates from the set CCZ, CZ, Z conjugated by Hadamards, or CZ, T conjugated by Hadamards, cannot be weakly classically simulated to constant multiplicative error unless PH collapses. In 2016 Bremner Shepherd and Montanaro [74] improved this to a hardness of simulation to constant additive error, under an additional hardness-on-average assumption. This additive hardness was later improved to cover lower depth circuits (namely $\tilde{O}(n^{1/2})$ depth, vs. $O(n)$ depth in their early construction) with nearest-neighbor interactions on a 2D grid under a different hardness-on-average assumption [73]. Note that [74, 73] proved their corresponding anticoncentration theorems. In Chapter 4 we will extend the hardness of constant multiplicative error to a much broader family of commuting operations [66].

Modified Clifford Circuits: Quantum circuits beginning in the computational basis any applying polynomially many Clifford gates (i.e. gates from the set CNOT, H, P) are efficiently classically simulable (in both the strong and weak sense) by the Gottesman Knill Theorem [133]. However several authors have shown that modified forms of Clifford circuits are hard to simulate classically. Jozsa and Van den Nest [163] showed that Clifford circuits with some $|0\rangle$ inputs, and some magic state inputs, are not weakly simulable to constant

multiplicative error unless PH collapses. In Chapter 5, we will introduce a new modified version of Clifford Circuits, in which the Clifford gates are conjugated by a one-qubit unitary U on every qubit. We show that essentially any non-Clifford U allows one to perform sampling tasks which are hard to multiplicatively simulate. We also show that such circuits are hard to simulate to constant additive error under an additional hardness-on-average assumption [65]. Note that we are able to prove our corresponding anticoncentration theorem.

One clean qubit (DQC_1): In this model, one considers the power of quantum circuits with a universal gate set, but in which some of the inputs are in an unknown state. Specifically, in this model the first qubit is in the state $|0\rangle$, but the remaining $n-1$ qubits are in a uniformly random computational basis state unknown to the experimenter. (This is known as the "maximally mixed state"). This model was first introduced by Knill and LaFlamme [171], where they showed that this model can efficiently estimate the trace of an exponentially large unitary matrix (specified by a poly-sized quantum circuit). There is no known classical algorithm for this task. Shor and Jordan [222] showed that this model can estimate the Jones polynomial of the trace closure of a braid (whereas the Platt closure is BQP-complete [31, 176]). It in unclear of this model is capable of performing even universal classical computation, however it is known that one can compute log-depth classical circuits in DQC_1 [37]. In terms of sampling problems, Morimae, Fujii and Fitzsimons showed that DQC_1 circuits cannot be weakly simulated to constant multiplicative error unless PH collapses [195]. This was improved to constant additive error by Morimae [194] under additional assumptions; however the correct distribution of unitaries for additive hardness is left open in this model. In this model the corresponding anticoncentration theorem can be proven, essentially because the computations are so highly mixed.

Yang-Baxter Ball Permuting: In this model, one considers a system of n distinguishable particles on a line, each with their own momenta, which scatter off one another with a soft interaction. These particles obey what is known as the "Yang-Baxter" equation. Despite the simplicity of this model, in the presence of intermediate measurements, we show that it cannot be weakly simulated to constant multiplicative error unless PH collapses. We will discuss this model in Chapter 6 and in more detail in the full version of this work [14].

Permutational Quantum Computing: In this model, defined by Jordan, [159] one imagines that one can merely permute spin 1/2 particles and couple their angular momenta. This is similar to topological quantum computation, except the particles only care about their permutation and not the topology of the braid used to move them to that permutation. Jordan showed that this model can approximate matrix elements of irreducible representations of the symmetric group, which are not known to be tractable by classical devices. It remains open if the probability distributions output in this model are hard to weakly simulate.

Quantum Approximate Optimization Algorithm (QAOA): In 2016 Farhi and Harrow showed that QAOA quantum circuits cannot be weakly classically simulated to constant multiplicative error [111]. Their model consists of alternating between applying a local problem Hamiltonian, diagonal in the Z basis, and the Hamiltonian $\sum_i X_i$ to the equal superposition input state. They show that even one alternation suffices for this hardness result. They leave the question of additive hardness open.

We next list those results which show hardness of sampling using a universal set of quantum gates.

Fourier Sampling: Fefferman and Umans proposed demonstrating quantum advantage through Fourier Sampling. In this model, one samples an output y with probability

proportional to some the evaluation of some"efficiently specifiable" polynomial on y. This can be achieved quantumly using the Quantum Fourier Transform. Fefferman and Umans show hardness of sampling in this model to constant additive error under two additional complexity theoretic assumptions, one of which is an anti-concentration conjecture, and one which is a hardness-on-average conjecture [114]. Their conjectures have the advantage that they are each strict weakening of the corresponding conjectures used to show advantage for Boson Sampling [11]. However unlike Boson Sampling, their model requires a universal gate set to implement.

Random quantum circuits: Several authors have proposed using random universal circuits to demonstrate quantum advantage. Boixo *et al.* [61] proposed using random quantum circuits on a 1D or 2D lattice. They showed hardness of sampling to constant additive error assuming non-collapse of PH, as well as an anticoncentration conjecture (for which they give numerical evidence) and a hardness-on-average conjecture. Aaronson and Chen [15] also proposed using random quantum circuits applied to a 2D grid of qubits. They show quantum advantage for such circuits using a different type of complexity assumption - namely that it is difficult for a classical computer to predict which outputs of a random quantum circuit occur with relatively high probability compared to the other possible outputs.

2.2.5 Restricted gate set conjecture

In the last section, we saw a number of examples of restricted quantum gate sets which despite being non-universal, nevertheless can perform sampling problems which are hard to weakly simulate to constant multiplicative error.

On the other hand, there are a small number of quantum gate sets for which we do have an efficient classical simulation algorithm. For instance, Clifford gates - i.e. those gates from the set CNOT, H, and P (phase by i) - can be efficiently classically simulated by the Gottesman-Knill Theorem [133]. These gates perform a discrete set of unitaries on any number of qubits.

A natural goal is to fully classify the computational power of all possible restricted gate sets. That is, for any gate set G, determine the computational complexity of simulating poly-sized circuits consisting of gates from G - call this model $\mathcal{C}_G$. Thus far, all examples of gate sets which we have studied fall into one of three categories. Either

- Circuits from $\mathcal{C}_G$ are efficiently classically simulable. For example the Clifford group has this property.
- Postselected circuits from $\mathcal{C}_G$ are capable of PostBQP, and therefore $\mathcal{C}_G$ circuits can perform sampling problems which are hard to weakly simulate to constant multiplicative error by Lemma 2.2.2. For instance commuting circuits have this property.
- Circuits from $\mathcal{C}_G$ are universal.

In the above, when we say a that the circuits from $\mathcal{C}_G$ are universal, this sometimes can occur in an encoded sense. This is known as *computational universality*, and this can occur even if the gate set is not physically universal (i.e. densely generates all unitaries). For example, if one has only real gates at one's disposal, then one can simulate complex gates on n qubits by simply adding an extra qubit to represent the real and complex amplitudes separately. There also exist more complicated encoded universality constructions. For instance, one can define a logical $|0\rangle_L$ and logical $|1\rangle_L$ state over $k > 1$ qubits, and show that

the gates in $\mathcal{C}_G$ can perform BQP circuits over the logical subspace. For examples of such constructions see [103, 77].

It is therefore natural to conjecture that these are the only three possibilities. We call this the Gate Set Trichotomy Conjecture, which we describe below:

Conjecture 2.2.1 (Gate set trichotomy conjecture). *For any two-qubit gate set G, poly-time quantum computing with G starting in the computational basis is either*

1. *Universal for* BQP
2. *Universal for* PostBQP *under postselection, and therefore capable of performing sampling problems which are hard to (exactly) weakly simulate by classical computers unless* PH *collapses by Lemma 2.2.2.*
3. *Efficiently weakly classically simulable*

Here "universality" refers to computational universality. As 1 implies 2 (so long as G admits a Solovay-Kitaev algorithm, for example if it is closed under inversion), the conjecture states that sampling hardness is pervasive - all gate sets can perform hard sampling problems, except for those which are efficiently classically simulable. The conjecture also states that there are no further intermediate degrees of computational hardness, other than sampling hardness unless PH collapses. We believe proving this Conjecture is an important open problem, as the proof would likely require classifying and describing those gate sets which fall into categories 2 and 3. This would provide a "complete map" of the space below BQP defined by restricted gate sets.

Proving this Conjecture, on the other hands, seems to be a very difficult task, for several reasons.

First, several simpler versions of the conjecture remain open. For example, it is open to fully classify which sets of two-qubit gates generate all unitaries. Although Lloyd [180] and others [102, 240, 90, 52] have shown that a Haar-random two-qubit gate is universal with probability 1, we haven't yet classified what happens on the remaining set of measure zero. The closest related result by Childs, Leung, Mančinska, and Ozols [90] classifies the set of two-qubit Hamiltonians which give rise to $SU(4)$ when acting on two qubits, and are hence physically universal. However, it even remains open to classify which two-qubit Hamiltonians are physically universal on three or more qubits! Answering this question for gates will be an even more difficult task, since as we discussed in Section 2.1.4, proving a discrete get set generates an infinite group can be challenging.

Second, the power of quantum gate sets can change under conjugation by a unitary. For example, gates which are diagonal in the computational basis are trivially efficiently classically simulable, since they do nothing to a basis input state. However, Bremner Jozsa and Shepherd showed that conjugating such circuits by Hadamards can make such circuits hard to simulate classically [72]. This means that is is difficult to use algebraic results to prove the conjecture. For instance, although there is in principle a classification of all possible finite subgroups of $SU(4)$ [142], this classification is only performed up to conjugation by a unitary. So in order to use this approach to solve the conjecture, one would need to classify the power of all representations of these subgroups conjugated by any possible unitary. This would be a challenging and tedious task. Even if this were possible, it could be that some gate sets are not universal when applied to two qubits, but do become universal when applied to three qubits ([90] gave an example of this in the Hamiltonian setting). Therefore

one would then need to use the classification of finite subgroups of $SU(8)$ to continue - but this classification remains open! So this algebraic approach does not seem a viable approach to completing the classification.

Third, as discussed in Section 2.2.3 even if a gate set is physically universal under post-selection, in order to prove sampling hardness one must show that the gadgets used obey a Solovay-Kitaev Theorem to apply Lemma 2.2.3. Therefore, one must either find inverses of the postselection gadgets applied, find gadgets to perform the Weyl Group or Clifford Group, or prove an inverse-free Solovay-Kitaev Theorem to show sampling hardness.

Fourth, there are very few examples of gate sets which remain non-universal on arbitrary numbers of qubits. To our knowledge, the only examples of 2-qubit gate sets which are *NOT* known to be physically universal on arbitrary numbers of qubits are:

- Real gates - which are computationally universal.
- Clifford gates - i.e. the gate set CNOT, Hadamard, Phase by i.
- Diagonal gates
- Gates which preserve the parity of the input strings, such as gates arising from the exchange interaction [103] or from matchgates [236, 229, 77].
- Subsets of the above gate sets, and conjugations of the above gate sets by the same unitary on every qubit.

Even going through all of these examples, and considering their power under taking subsets and conjugating by unitaries, would be very time consuming. For instance the Clifford group contains 57 different subgroups [137]. Even with this, going through each possible conjugation of each possible subgroup would be a tedious task.

In short, proving Conjecture 2.2.1 seems to be beyond the reach of our current techniques. However, we still believe it is an important goal to work towards, as it can help to provide a more complete map of the space below BQP, and may lead to new models of sampling problems to demonstrate quantum advantage. Several of the results of this thesis can be seen as making progress towards this conjecture. Chapter 3 completely classifies the power of restricted two-mode optical gates in the linear optics setting [63]. Chapter 4 completely classifies the power of two-qubit commuting Hamiltonians, and shows they are either efficiently simulable or else produce hardness of sampling results [66]. Chapter 5 completely classifies the power of Clifford circuits conjugated by the same one-qubit unitary U on each qubit, showing essentially that for any U which is non-Clifford, such conjugated Clifford circuits give rise to hard sampling problems [65]. We hope these works will stimulate further exploration of Conjecture 2.2.1 and help lead to the discovery of more weak models of quantum computation.

2.3 Representation Theory & Lie Algebras

In many of our proofs, we will make use of representation theory and Lie algebras. In general, the set of operations densely generated by a gate set G forms a group - and the set of matrices generated is also a particular *representation* of that group, i.e. an embedding of the group into matrices. Representation theory places constraints on what these embeddings can look like, and therefore is quite useful when studying the power of restricted quantum gate sets. If the group generated is continuous, then the group generated is also a Lie group.

In this case a "linearized" version of the group, known as the Lie algebra, is often easier to deal with when proving universality results. In this section we'll briefly review the basics of representation theory and Lie algebras; for a more thorough introduction see [141, 76, 158]. Part of this review also appears in my paper with Scott Aaronson, Greg Kuperberg and Saeed Mehraban [14].

2.3.1 Representations, Irreps, Character Tables

Representation theory is the study of embeddings of groups into matrices (or more generally linear maps over abstract vector spaces). Specifically, a representation ρ of a group G is a homomorphism from G to the group of isomorphisms of a linear space : $G \to GL(V, \mathbb{C})$, for some vector space V. Let g be any element of G, with its inverse g^{-1}, and e and 1 as the identity elements of G and $GL(V, \mathbb{C})$, respectively. Given the definition, $\rho(g^{-1}) = \rho(g)^{-1}$, and $\rho(e) = 1$ are immediate. In this work we will always have $V = \mathbb{C}^d$, i.e. we will always work over a finite dimensional complex vector space, as we will assume G is a finite group as well. Of course one can define representations of infinite groups as well on finite or infinite dimensional vector spaces, but as we won't use these for our results, we refer the interested reader to [76].

There are several representations which exist for all groups. The first is the trivial representation - in which $\rho : G \to \{I \in GL(V, \mathbb{C})\}$, i.e. every element of the group is represented by the identity. The second is the regular representation. Here the group G is represented over the vector space $\mathbb{C}^{|G|}$, where each basis vector e_g corresponds to a group element $g \in G$. Each group element is represented by a permutation matrix which describes what permutation g applies to the elements of G (each of which is a basis vector).

A representation ρ on V is called an irreducible representation (irrep), if it has no stable subspaces other than 0 and V. For example, the regular representation of the cyclic group on 3 elements is NOT an irrep - as it preserves the vector $e_1 + e_2 + e_3$. Irreps are the basic building blocks of all representations - i.e. any representation of any group can be decomposed into irreps by a suitable change of basis. More formally, for any reducible representation ρ, there exists a change of basis U such that conjugating the matrices of ρ renders them block-diagonal, where each block corresponds to an irrep [13]. In this case we write that

$$\rho \cong \bigoplus_i \rho_i$$

Sometimes this is also written just in terms of the vector spaces involved, i.e. the representation is implicit:

$$V \cong \bigoplus_i V_i$$

A fascinating fact of representation theory is that for any finite group, there are only a finite number of irreps! In fact the number of irreps is equal to the number of conjugacy classes in the group. The representations are further constrained by the fact that the sums of the dimensions squared of the irreps must equal the order of the group. In short, the representations of finite groups are highly constrained - up to a change of basis, they must consist of a direct sum of one of a small number of possible irreps.

Note that if an irrep appears more than once in a reducible representation, then the structure of the representation can be slightly more complicated. In this case, the number

[13] Beware this is not necessarily of infinite groups acting on infinite-dimensional vector spaces!

of times a particular irrep V_i appears in the representation is known as the multiplicity of the irrep - denote this m_i. These isomorphic subspaces can be grouped together to $V \cong m_1V_1 \oplus m_2V_2 \oplus \ldots m_kV_k$. Then $\dim V = \sum_j m_j \dim V_j$. The structure of such a decomposition is isomorphic to $\bigoplus_j V_j \otimes X_j$, where X_j is the multiplicity space of V_j and is a vector space of dimension m_j. Decomposition of a representation onto the irreducible ones is unique up to isomorphism and multiplicities and the dimensionality of the irreps do not depend on the decomposition.

Information about representations of a finite group is often summarized in a character table. In the context of finite-dimensional representations of finite groups, the character table is a square matrix. Each row is an irrep of the group, and each column is a conjugacy class of the group. The entry of the character tables shows the trace of an element of that conjugacy class under that particular irrep. (Traces are the same within each conjugacy class by the cyclic property of trace).

Finally, we will also introduce the notion of a dual representation. A dual representation of G is a homomorphism from G into the group of linear maps : $V \to \mathbb{C}$. This is called the dual space $V^\star$, and V is viewed as the space of column vectors, then its dual space is a row space. For any vector spaces V and W, the two can be combined into a larger linear structure, $V \otimes W^\star$, as the set of linear maps from W to V. Let M_1 and M_2 be two elements of $GL(V, \mathbb{C})$ and $GL(W, \mathbb{C})$, respectively. Then, viewing $V \otimes W^\star$ as a vector space, the object (M_1, M_2) acts on $x \in V \otimes W^\star$ with $M_1 x M_2^{-1}$. Then, if M_1 and M_2 are two representations of G on V and W, then (M_1, M_2) is a representation of G on $V \otimes W^\star$, as a vector space. Notice that the inverse on M_2 is needed in order to have (M_1, M_2) act as a homomorphism. The dual representation M of V is then the representation on $\mathbb{C} \otimes V^\star$, when $M_2 = M$, and M_1 is the one dimensional trivial representation. This is just saying that the dual representation $M^\star$ of M on $V^\star$, maps $\langle\psi|$ to $\langle\psi|M(g^{-1})$, if we view the dual space as the usual row space. If we define an inner product as the action of the dual of a vector on itself, then G, as a representation, sends orthonormal basis to orthonormal basis. This implies that every representation of a finite group is isomorphic to a unitary representation. That is, any non-unitary representation becomes unitary after a change of basis. Let M be a representation on V. Then, we say $W \subseteq V$ is called stable under M, if for any $x \in W$, $Mx \in W$. Then, M restricted to W is called a sub-representation.

2.3.2 Representation Theory of the Symmetric Group

Most of this review is also appears in my paper with Scott Aaronson, Greg Kuperberg and Saeed Mehraban [14] and is based on [158]. We are interested in two mathematical structures, the group algebra of the symmetric group $\mathbb{C}S_n$, and the unitary regular representation of the symmetric group. As it turns out, the two structures are closely related to each other, and also to the group generated by the ball permuting gates discussed in Chapter 6. A group algebra is an extension of a group to an algebra, by viewing the members of the group as linearly independent basis of a vector space over the field $\mathbb{C}$. Therefore, in addition to the group action an action of $\mathbb{C}$ on S_n is needed, by the map $(\alpha, \sigma) \mapsto \alpha S_n$, and also addition of vectors in the usual sense. Therefore, a group algebra consists of all elements that can ever be generated by vector on vector composition and linear combination of vectors over $\mathbb{C}$. Any element of $\mathbb{C}S_n$ can be uniquely written as $\sum_{\sigma \in S_n} \alpha_\sigma \sigma$, with $\mathbb{C}$ coefficients α_σ. If we

add a conjugation convolution † with maps $\sigma^\dagger = \sigma^{-1}$, and $\alpha^\dagger = \alpha^\star$, then for any element $v \in \mathbb{C}S_n$, $v^\dagger v = 0$, if and only if, $v = 0$. In order to see this, let $v = \sum_{\sigma \in S_n} \alpha_\sigma \sigma$. Then, $v^\dagger v = \sum_\sigma |\alpha_\sigma|^2 e + \ldots = 0$. A zero on the right hand side implies zeroth of all the vector components, including the component along e, which implies $\alpha_\sigma = 0$ for all $\sigma \in S_n$, and therefore $v = 0$. Let e be the identity element of S_n, consider an element $p \in \mathbb{C}S_n$ to be a projector if it has the property $p^2 = p$. Two projectors p and q are called orthogonal if $p.q = 0$. Then $(e-p)^2 = e - p$ is also a projector, and also $p(e-p) = 0$ are orthogonal projectors. 0 is trivially a projector. Therefore, the group algebra decomposes as

$$\mathbb{C}S_n = \mathbb{C}S_n e = \mathbb{C}S_n(e-p) + p = \mathbb{C}S_n(e-p) \oplus \mathbb{C}S_n p.$$

A projector is called minimal if it cannot be written as the sum of any two other projectors other than 0 and itself. Let p^μ be a list of minimal projectors summing $\sum_\mu p^\mu = e$, then the decomposition of the group algebra into minimal parts is according to

$$\mathbb{C}S_n = \bigoplus_\mu \mathbb{C}S_n p^\mu.$$

p^μ are known as Young symmetrizers, and we are going to mention them later.

The regular representation of S_n, also denoted by $\mathbb{C}S_n$, is the unitary representation of S_n onto the usual Hilbert space $\mathbb{C}S_n$ spanned by the orthonormal basis $\{|\sigma\rangle : \sigma \in S_n\}$. It is well known that for any regular representation, the dimension of each irrep is equal to the multiplicity of the irrep, and therefore $\mathbb{C}S_n$ decomposes into irreducible representations of the following form

$$\mathbb{C}S_n \cong \bigoplus_\lambda V_\lambda \otimes X_\lambda,$$

with $\dim X_\lambda = \dim V_\lambda =: m_\lambda$, and indeed $\sum_\lambda m_\lambda^2 = n!$. Here X_λ is again the multiplicity space, and V_λ corresponds to each irrep. It is tempting to make a connection between the group algebra and regular representation of the symmetric group. As described earlier, S_n can act on the Hilbert space $\mathbb{C}S_n$ in two ways; the left and right, $L, R : S_n \to U(\mathbb{C}S_n)$, unitary regular representation, with the maps $L(\sigma)|\tau\rangle = |\sigma \circ \tau\rangle$ and $R(\sigma)|\tau\rangle = |\tau \circ \sigma^{-1}\rangle$. Also, similar left and right structure can be added to the group algebra. Clearly, L and R representations commute, and it can be shown that the algebra generated by L is the entire commutant of the algebra generated by R. Putting everything together, inspired by the theory of decoherence free subspaces, and the defined structures, one can show that the left (A) and right (B) algebras and the Hilbert space $\mathbb{C}S_n$ decompose according to

$$A \cong \bigoplus_\lambda M(m_\lambda) \otimes I(m_\lambda),$$

$$B \cong \bigoplus_\lambda I(m_\lambda) \otimes M(m_\lambda),$$

and,

$$\mathbb{C}S \cong \bigoplus_{\lambda} V(m_\lambda) \otimes X(m_\lambda).$$

This is indeed a nice and symmetric structure. Indeed each irrep V_λ is an invariant subspace of the X operators, and it cannot be reduced further. It remains to demonstrate the structure of the irreps λ, and to study the action of X operators which exchange two elements on these subspaces.

The irreducible representations of the symmetric group S_n are indexed by the partitions of n. Remember that a partition of n is a sequence of non-ascending positive numbers $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \lambda_k$ summing to n, i.e., $\sum_j \lambda_j = n$. The number of partitions of n grows like $\exp \Theta(\sqrt{n})$. As described earlier, each partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ is related to a diagram, called the Young diagram, which consists of k horizontal rows of square boxes $r_1, r_2, \ldots, r_k$. The Young diagram is then created by paving the left-most box of r_1 to the left-most box of r_2, and so on. For a Young diagram λ, the dual diagram $\tilde{\lambda}$, is another Young diagram, whose rows are the columns of λ. A Young tableau t^λ with the shape λ, is a way of bijective assigning of the numbers in $[n]$ to the boxes of λ. We will use t^λ and simply t with the shape λ interchangeably. A permutation $\pi \in S_n$ can act on a Young tableau t^λ by just replacing the content of each box to the its image under π, i.e., if a box contains j, after the action of π it will be replaced with $\pi(j)$. A tableau is called standard, if the numbers in each row and column are all in ascending orders. The number of standard tableau for each partition of shape λ is denoted by f^λ.

Let t be a tableau with shape λ. Define $P(t)$ and $Q(t) \subseteq S_n$ to be sets of permutations that leave each row and column invariant, respectively. Then the projectors of the $\mathbb{C}S_n$ group algebra are according to the Young symmetrizers, one for each standard tableau

$$p^t = \frac{1}{f^\lambda} \sum_{\pi \in C(t)} \sum_{\sigma \in R(t)} sgn(\pi) \pi \circ \sigma.$$

These subspaces correspond to all of the irreducible invariant subspaces of S_n. The dimension for each of these subspaces is the number of standard tableaus of each partition, and it is computable using the hook lengths. The hook of each box in a partition of shape λ is consists of the box itself along with all boxes below and at the right of the box. The hook length of each box is the number of boxes contained in that hook, and the hook length h^λ of the shape λ is the multiplication of these numbers for each box. Then, the dimension of the irrep corresponding to λ is according to $f^\lambda = n!/h^\lambda$.

The Young-Yamanouchi Basis

In order to talk about quantum operations, an orthonormal basis for the discussed subspace is needed. It would be nice if we had a direct description of the basis, in a way that the action of X operators (which exchange two labels) on these subspaces is clear. Moreover, we seek an inductive structure for the orthonormal basis of the irreps that is adapted to the nested subgroups $S_1 \subset S_2 \subset \ldots \subset S_n$. By that we mean states that are marked with quantum numbers like $|j_1, j_2, j_3, \ldots, j_k\rangle$, such that while elements of S_n affect all the quantum numbers, for any $m_1 < n$, elements of S_n restricted to the first m_1 labels affects the first k_1 quantum numbers only, and act trivially on the rest of the labels. Also, for any $m_2 < m_1 < n$, the elements of S_n restricted to the first m_2 labels affect the first $j_2 < j_1 < k$

quantum numbers only, and so on.

Fortunately, such a bases exist, and are known as the subgroup adapted Young Yamanouchi (YY) bases [158]. These bases are both intuitive and easy to describe: for any partition of shape λ, mark an orthonormal basis with the standard Young Tableaus of shape. Agree on a lexicographic ordering of the standard tableaus, and denote these basis corresponding to the partition λ, by a $\{|\lambda_j\rangle\}_{j=1}^{f^\lambda}$. Denote the action of a swap (i,j) on $|\lambda_l\rangle$ by $|(i,j).\lambda_l\rangle$, to be the basis of a tableau that is resulted by exchanging location of i and j in the boxes. Suppose that for such tableau t, the number j (i) is located at the r_j and c_j (r_i and c_i) row and column of t, respectively. Then, define the axial distance d_{ij} of the label i from label j of on each tableau to be $(c_j - c_i) - (r_j - r_i)$. Or in other words, starting with the box containing i walk on the boxes to get to the box j. Whenever step up or right is taken add a -1, and whenever for a step down or left add a 1. Starting with the number 0, the resulting number in the end of the walk is the desired distance. Given this background, the action of $L_{(k,k+1)}$ on the state $|\lambda_i\rangle$, is according to

$$L_{(k,k+1)}|\lambda_i\rangle = \frac{1}{d_{k+1,k}}|\lambda_i\rangle + \sqrt{1-\frac{1}{d_{k+1,k}^2}}|(k,k+1).\lambda_i\rangle.$$

Three situations can occur: either k and $k+1$ are in the same column or row, or they are not. If they are in the same row, since the tableau is standard, k must come before $k+1$, then the axial distance is $d_{k+1,k} = 1$, and the action of $L_{(k,k+1)}$ is merely

$$L_{(k,k+1)}|\lambda_i\rangle = |\lambda_i\rangle.$$

If the numbers are not in the same column, k must appear right at the top of $k+1$, and the action is

$$L_{(k,k+1)}|\lambda_i\rangle = -|\lambda_i\rangle.$$

Finally, if neither of these happen, and the two labels are not in the same row or column, then the tableau is placed in the superposition of itself, and the tableau wherein k and $k+1$ are exchanged. Notice that if the tableau $|\lambda_i\rangle$ is standard the exchanged tableau $|\lambda_i\rangle$ is also standard. This can be verified by checking the columns and rows containing k and $k+1$. For example, in the row containing k, all the numbers at the left of k are less than k, then if we replace k with $k+1$, again all the numbers on the left of $k+1$ are still less than $k+1$. Similar tests for the different parts in the two rows and columns will verify $(k,k+1)\lambda_i$, as a standard tableau. The action of $L_{k,k+1}$ in this case is also an involution. This is obvious for the two cases where k and $k+1$ are in the same row or column. Also, in the third case if the action of $L_{(k,k+1)}$ maps $|\lambda\rangle$ to $\frac{1}{d}|\lambda\rangle + \sqrt{1-\frac{1}{d^2}}|t\circ\lambda\rangle$ then a second action maps $|t\circ\lambda\rangle$ to $\frac{-1}{d}|t\circ\lambda\rangle + \sqrt{1-\frac{1}{d^2}}|\lambda\rangle$, and therefore

$$L_{(k,k+1)}^2|\lambda\rangle = \frac{1}{d}(\frac{1}{d}|\lambda\rangle + \sqrt{1-\frac{1}{d^2}}|t\circ\lambda\rangle) + \sqrt{1-\frac{1}{d^2}}(\frac{-1}{d}|t\circ\lambda\rangle + \sqrt{1-\frac{1}{d^2}}|\lambda\rangle) = |\lambda\rangle.$$

Given this description of the invariant subspaces, we wish to provide a partial classification of the image of the ball permuting gates on each of these irreps. The hope is to find

denseness in $\prod_{\lambda} SU(V_\lambda)$, on each of the irreps V_λ, with an independent action on each block. In this setting, two blocks λ and μ are called dependent, if the action on λ is a function of the action on μ, i.e., the action on the joint block $V_\lambda \oplus V_\mu$ resembles $U \times f(U)$, for some function f. Then, independence is translated to decoupled actions like $I \times U$ and $U \times I$.

Throughout, the $\lambda \vdash n$, means that λ is a partition of n. We say $\mu \vdash n+1$ is constructible by $\lambda \vdash n$, if there is a way of adding a box to λ to get μ. We say a partition $\mu \vdash m$ is contained in $\lambda \vdash n$, for $m < n$, if there is a sequence of partitions $\mu_1 \vdash m+1$, $\mu_2 \vdash m+2, \ldots, \mu_{n-m-1} \vdash n-1$, such that μ_1 is constructible by μ, λ is constructible by μ_{n-m-1}, and finally for each $j \in [n-m-2]$, μ_{j+1} is constructible by μ_j. We also call μ a sub-partition of λ. A box in a partition λ is called removable, if by removing the box the resulting structure is still a partition. Also, define a box to be addable if by adding the box the resulting structure is a partition.

Theorem 2.3.1. *The Young-Yamanouchi bases for partitions of n are adapted to the chain of subgroups $\{e\} = S_1 \subset S_2 \subset \ldots \subset S_n$.*

Proof. Let $\lambda \vdash n$, and t be any standard tableau of shape λ. We construct some enumeration of states in the Young-Yamanouchi basis of λ which is adapted to the action of subgroups. For any $m < n$, since t is a standard tableau, the numbers $1, 2, 3, \ldots, m$, are all contained in a sub-partition $\mu \vdash m$ of λ. This must be true, since otherwise the locus of numbers $1, 2, 3, \ldots, m$ do not shape as a sub-partition of λ. Let ν be the smallest sub-partition of n that contains these numbers. Clearly, $|\nu| > m$. The pigeonhole principle implies that, there is a number $k > m$ contained somewhere in ν. The box containing k is not removable from ν, since otherwise you can just remove it to obtain a sub-partition smaller than ν that contains all of the numbers in $[m]$. Therefore, if k is in the bulk of ν, then both the row and column containing k are not in the standard order. If k is on a vertical (horizontal) boundary, then the column (row) of the box containing k is not standard.

Let λ_k be the smallest sub-partition of λ that contains $[k]$. Then the enumeration of the basis is according to $|\lambda_1, \lambda_2, \ldots, \lambda_n\rangle$. Here, $\lambda_n = \lambda$, and λ_1 is a single box. From before, for any $j < n$, λ_{j+1} is constructible by λ_j. For $m < n$, let S_m be the subgroup of S_n, that stabilizes the numbers $m+1, m+2, \ldots, n$. For any $k \leq m$, $L_{(k,k+1)}$ just exchanges the content of boxes within λ_m, and therefore leaves the quantum numbers $\lambda_{m+1}, \lambda_{m+2}, \ldots, \lambda_n$ invariant. Moreover, the box containing m is somewhere among the removable boxes of λ_m, since otherwise, as described in the last paragraph, the tableau λ_m is not standard. The box containing $m-1$ is either right above or on the left side of m, or it is also a removable box. In the first two cases, the action of $L_{(m-1,m)}$ is diagonal, and the quantum numbers are intact. In the third case, the only quantum numbers that are changed are λ_{m-1} and λ_m. □

Consider now the action of S_{n-1} on an element $|\lambda_1, \lambda_2, \ldots, \lambda_n = \lambda\rangle$. In any case λ is constructible by λ_{n-1}, and the construction is by adding an addable box to λ_{n-1}. In other words, λ_{n-1} can be any partition $\vdash n-1$, that is obtained by removing a removable box from λ. These observations, all together, lead to a neat tool:

Lemma 2.3.1. *(Branching.) Under the action of S_{n-1}, $V_\lambda \cong \bigoplus_{\substack{\mu \vdash n-1 \\ \mu \subset \lambda}} V_\mu$.*

Proof. The proof is directly based on the structure of the YY bases. What we would like to emphasize here is that the multiplicity free branching rule of the symmetric group is manifest in the structure of the YY bases. For other formal proofs see [158].

Choose an orthonormal basis according to YY. Enumerate the removable boxes of λ by $1, 2, \ldots, p$. Clearly, in any standard tableau of λ, the box containing n is a removable one. Group the tableaus according to the location of n. Clearly, each subspace corresponds to a partition $\mu \vdash n-1 \subset \lambda$. Call these partitions $\mu_1, \mu_2, \ldots, \mu_p$, according to the enumeration of removable boxes. Also denote the space V_{μ_j} correspondingly. For any μ_j, any element of S_{n-1}, acted on V_{μ_j}, generates a vector within V_{μ_j}. In other words, these subspaces are stable under S_{n-1}. □

2.3.3 A brief overview of Lie groups and Lie algebras

Here we quickly review the pertinent facts about Lie algebras used in our results; for a more thorough introduction see [141]. In fact here we will only use facts about matrix groups and algebras, as it is all we will need for our proofs.

A Lie group is a group of matrices which have a continuous structure (i.e. which are also a smooth manifold). For instance, the set of all unitary matrices is a Lie group, as is the set of all orthogonal matrices. A Lie algebra is a structure defined to be the "tangent space" to a Lie group G at the Identity matrix. More formally, let $\gamma(t)$ be a one-parameter family of matrices such that $\gamma(t) \in G$ for all $t \in \mathbb{R}$, and such that $\gamma(0) = I$. Then clearly one can consider taking the derivative of this matrix and evaluating it at zero. This matrix $g = \frac{d}{dt}\gamma(t)|_{t=0}$ belongs to the tangent space of the group at the identity, i.e. the Lie algebra of G, denoted $\mathfrak{g}$. It turns out that from the simple group axioms, one can derive that $\mathfrak{g}$ is closed under real scalar multiplication (which follows from rescaling $t \to ct$), addition (which follows from G being closed under multiplication), and commutator, i.e. if $g, h \in \mathfrak{g}$ then $[g, h] = gh - hg \in \mathfrak{g}$ (which follows from G being closed under conjugation). In short a Lie algebra is a vector space associated with the group which is also closed under commutators. For instance, the Lie algebra corresponding to the unitary group $SU(n)$, denoted $\mathfrak{su}(n)$, consists of all skew-Hermitian matrices. The Lie algebra corresponding to the orthogonal group $SO(n)$, denoted $\mathfrak{so}(n)$, consists of all skew-symmetric matrices.

While one can move from the Lie Group to the Lie algebra by taking derivatives, to move from the algebra to the group, one uses the exponential map, i.e.

$$\exp(A) = I + A + \frac{A^2}{2} + \frac{A^3}{6} + \ldots + \frac{A^i}{i!} + \ldots$$

The exponential map takes an element of the Lie algebra and maps it into the group G. In many cases, this map is *onto* the group G as well - i.e. all elements of the group can be obtained by exponentiating an element of the algebra. This occurs if the group G is closed (i.e. closed under limits) and simply connected. For instance this property is true of $U(n)$ - i.e. any unitary can be obtained by exponentiating some skew-Hermitian matrix (i.e. a Hamiltonian). In these cases the algebra $\mathfrak{g}$ generated fully characterizes the group.

This fact can be very useful in proving universality results. Suppose that one knows a set of unitary gates densely generate a continuous group G, and suppose one can show that some particular $g, h \in \mathfrak{g}$. Then, if one can show that g and h generate all of $\mathfrak{su}(n)$ when taking their closure under addition, scalar multiplication and commutators, then one has shown that $\mathfrak{g} = \mathfrak{su}(n)$. This immediately implies that $G = SU(n)$, so the gate set is

universal. This style of argument will be used in Chapter 4, and also appears many places in quantum information, e.g. [180, 90].

2.4 Preliminaries above BQP

In this section we will briefly discuss what is known about the space "above" BQP from the perspective of computational complexity .

2.4.1 Previously considered modifications to quantum theory

Several modifications to quantum theory have previously been studied from the perspective of computational complexity. Here we will briefly review the modifications previously considered. Most of the modifications lead to extremely powerful models of computation; however it remains open if these powerful models of computation can be made robust to noise, as they often make use of exponentially precise features of the maps applied.

Nonlinearities: In this model, one imagines that in addition to standard quantum operations like unitaries and measurements, one also has the ability to apply a nonlinear map M to the quantum state. Abrams and Lloyd [22] showed that essentially any such map should allow one to solve NP-hard and even PP-hard problems in polynomial time. Their algorithm works by the fact that any nonlinear map must "stretch" some area of the Bloch sphere; one can then use this to "pry apart" exponentially close quantum states. We will make this more precise in Chapter 9 by showing rigorously that any smooth diffeomorphism of the Bloch sphere admits such an algorithm, as well as exploring connections between this model and superluminal signaling. Aaronson [6] also studied a version of this model, and showed that if the set of all invertible nonlinearities M are allowed, then the power of the resulting class is equivalent to PP. (In contrast, Abrams and Lloyd only considered adding a single nonlinearity M, and argued this generically contains PP.). Note that in general the power of nonlinear quantum mechanics can depend subtly on the way the nonlinearity acts on mixed states [58].

Modifications to the Born rule: Aaronson [6] studied the power of quantum theory with a modified Born rule. In particular, he considered the power of quantum theory in which evolution is by unitary gates as usual, but measurement statistics are proportional to the pth power of the amplitude. He showed that this class contains PP for all $p \neq 2$, and is furthermore contained in PP if p is an even integer. We will revisit this modification, and explore its relation to superluminal signaling, in Chapter 9.

Closed Timelike Curves: Several authors have considered the power of quantum computing in the presence of Deutschian Closed Timelike Curves (CTCs) [101]. In this model, the closed timelike curve doesn't allow one to send arbitrary information back into the past, but rather imposes a consistency condition on the CTC to avoid the grandfather paradox. Mathematically the CTC is described as a fixed point of some operator. In the early 2002, Brun [78] observed that such a model would allow one to solve NP-hard problems efficiently. In 2004, Bacon [43] observed that such a model would allow one to solve PP-hard problems efficiently, even if the CTC only involved a single qubit, and furthermore one could make these CTC computations robust to small errors. In 2009, Aaronson and Watrous [20] showed that this model is exactly equivalent to PSPACE - that is this model can compute arbitrary PSPACE-complete problems and be simulated in PSPACE. They also showed that classical computation with CTCs is equivalent to PSPACE as well. Also note that in 2009, Bennett, Leung, Smith and Smolin [58] questioned whether the model used by the prior

authors is the correct one, in particular how the model ought to act on mixed states, and therefore cast doubt on the claim that CTCs would allow the solution of difficult problems.

Non-linear Schr$\acute{o}$dinger Evolution incorporated into dynamics: The prior models of modified quantum theory have considered the power of quantum mechanics supplemented by some nonlinear or non-quantum operation. In these models the nonlinearity is somehow separate from the normal evolution according to the Schr$\acute{o}$dinger equation, and can be turned on or off at will. Several authors have considered the power of nonlinear quantum theory, in which all evolution is required to be non-linear - in other words, the state evolves the entire time by some modification of the Schr$\acute{o}$dinger equation, and the nonlinearity can never be turned off. For instance, Meyers and Wong [189] considered the power of quantum mechanics under the Gross-Pitaevskii equation, a nonlinear version of the Schr$\acute{o}$dinger equation which nevertheless preserves the norm of the state. They showed that this model could allow for the efficient solution of NP-hard problems if one assumes the ability to perform exponentially precise Hamiltonian evolution. However they argue that if the precision of the evolution is limited, then the model would merely give a quadratic speedup over Grover search, and therefore could not realistically be used to solve NP-hard problems. They later generalized these results to more general forms of nonlinear evolution [190, 191]. Childs and Young [91] later improved Meyers and Wong's algorithm for the Gross-Pitaevskii equation to search in logarithmic time without resorting to exponentially precise Hamiltonian evolution, thus showing that such nonlinearities would indeed allow for the efficient solution of NP-hard and PP-hard problems.

2.4.2 A brief introduction to firewalls

In this section, be provide a brief overview of the firewalls paradox, aimed at a quantum information audience. For a more detailed description of the paradox see [35, 68], and for a nice introduction to black hole physics, see [149].

The firewalls paradox is a continuation of a longstanding debate over how quantum information behaves in the presence of black holes, known as the black hole information paradox. It reveals a fundamental inconsistency in our understanding of how general relativity and quantum theory interact with one another. This is a serious challenge to quantum gravity theorists and string theorists, who are seeking a unified theory to explain both phenomena.

In classical general relativity, a black hole is a singularity in space time. The black hole is surrounded by an event horizon - which signifies the point of no return for hitting the singularity - not even light can escape from the event horizon. The event horizon has no physical significance, as an observer falling into the black hole would not be able to tell when they crossed the horizon. The black hole is a one-way sink of matter and information - a drain at the end of the universe.

This picture radically changed in 1975, when Hawking showed that black holes not only absorb matter, but also emit matter [153]. In particular, he showed that black holes should emit thermal radiation, using calculations as to how quantum field theory should behave in the vicinity of the event horizon. This means that black holes would (very slowly) re-emit their mass back into the universe, and eventually disappear once all of its mass is radiated away.

However, Hawking realized shortly thereafter [152] that this thermal radiation seems to be in conflict with quantum mechanics. Quantum mechanics is reversible - in principle the evolution can be ran backwards - and therefore preserves information. However, the Hawking radiation was found to be thermal - and therefore does not contain any of the information

present in the material that formed the black hole. This seemed to suggest that black hole formation and evaporation is a non-unitary process, and therefore this process does not have a quantum mechanical description (at least in terms of unitary quantum mechanics). This would be a serious obstacle to constructing a quantum theory of gravity.

Others suggested that this problem might be resolved by assuming the Hawking radiation isn't perfectly thermal, but carries some small amount of information out of the black hole. But this suggestion led to a further problem known as the "Xeroxing paradox." Quantum mechanics does not allow for the cloning of quantum information - in other words there is no unitary map which takes as input the state $|\psi\rangle$ and outputs the state $|\psi\rangle \otimes |\psi\rangle$. However, if one naively considers that Hawking radiation is re-emitting the information inside which fell into black hole, then one finds that there are two copies of the same information existing at any given point in time - one inside the black hole (the original infalling matter, on its way to hitting the singularity), and one outside the black hole (the outgoing Hawking radiation). In a certain reference frame, both states are present in the same space-like slice - so the information is actually present in two places at the same time. This seems to contradict the no-cloning theorem. This observation led a long and complicated debate over how to reconcile conservation of information, the no-cloning theorem, and black hole physics (see [226] for a popular account).

One proposal that gained favor in the 90s was complementarity, put forth by Susskind, Thorlacius and Uglum [227]. In this picture, the interior and exterior of the black hole are "dual descriptions" of one another. This view is supported by the fact that the amount of information contained in a black hole is not given by its volume, but rather by its surface area [55]. The "complementarity" proposal essentially held that the two copies of the information inside and outside the black hole were two descriptions of the same information. A single observer could only see one copy or the other, so the information had not been cloned in an operational sense.

This picture was challenged in 2012 by Almheiri, Marolf, Polchinski and Sully [35]. They showed that there is an experiment that one can perform which contradicts the laws of quantum theory. In particular, by jumping into a black hole after half of its mass has been radiated away by Hawking radiation, one can observe a violation of monogamy of entanglement - the statement that three systems cannot be simultaneously maximally entangled with one another. Their violation of quantum theory occurs while merely making three assumptions:

1. "No drama" - an observer crossing the event horizon sees nothing special happen. More precisely, "effective field theory" is valid at the horizon.

2. Unitarity - black hole dynamics are reversible and unitary.

3. Scrambling - Hawking radiation re-emitted from the black hole is highly scrambled.

The paradox essentially follows from the following argument. By the first assumption, Hawking radiation is emitted from the black hole. This means there is a high amount of entanglement between the radiation leaving the black hole (A) and the radiation partners on the interior of the horizon. At the same time, the radiation leaving the black hole (A) must be highly entangled with some subsystem of the old Hawking radiation (R) emitted in the first part of the BH lifetime. (This is a generic feature of scrambling). Therefore we have a violation of monogamy of entanglement, which is apparent to an observer who collects the old radiation R, jumps into the black hole and catches the ingoing and outgoing Hawking modes A and B.

There have been a number of proposed resolutions to this new paradox. The authors of [35] proposed that one should give up on the "no drama" condition, and instead proposed that there is a "firewall" (a wall of high-energy radiation) at the horizon which prevents observers from crossing the horizon. Others such as Unruh have suggested to give up on the unitarity of black hole evolution, instead allowing for black holes to be a one-way sink of quantum information. Bousso suggested that perhaps the ingoing Hawking modes B and the old radiation R are actually the same physical system, so there is no violation of monogamy [67] - this is an "extreme" version of complementarity. Harlow and Hayden [150] argued that the experiment would be too difficult to perform from the perspective of computational complexity, so the paradox could never be observed. In joint work with Ning Bao, Aidan Chatwin-Davies, Jason Pollack, and Henry Yuen, we have suggested the experiment may be information-theoretically infeasible due to scrambling of the interior of the horizon [47]. Other solutions include "remnants" - the black hole never fully evaporates - and changing the scrambling assumption. There have also been proposals to modify quantum theory to evade the paradox, such as the final-state projection model of Horowitz and Maldecena [155, 181] and the state-dependence model of Papadodimas and Raju [203, 147].

Overall, at present it seems the paradox is far from being resolved. However, the paradox has led to an intense burst of activity in the quantum gravity community, as well as a renewed interest in modified theories of quantum mechanics. Regardless of its resolution, the black hole information paradox is a substantial challenge to anyone seeking to describe a quantum theory of gravity.

Part I

The space below BQP

Chapter 3

Classifying Beamsplitters

In this chapter, we begin exploring the space "below" BQP by studying linear optics. We consider a restricted model of quantum optics in which the set of allowed optical transformations is restricted. In particular, we imagine one only has access to single two-mode gate (known as a beamsplitter), which one can apply to any pair of modes. We completely classify the complexity of such models, and show that they either give rise to the full power of linear optics, or else give rise to trivially efficiently classically simulable models. One can see this result as a version of Conjecture 2.2.1 in the linear optical setting. The proof makes extensive use of representation theory.

This chapter is based on joint work with Scott Aaronson [63].

3.1 Introduction

In quantum optics, the Hilbert space is not built up as a tensor product of qubits; instead it's built up as a direct *sum* of optical modes. An optical *gate* is then just a unitary transformation that acts nontrivially on $O(1)$ of the modes, and as the identity on the rest. Whenever we have a k-mode gate, we assume that we can apply it to any subset of k modes (in any order), as often as desired. The most common optical gates considered are *beamsplitters*, which act on two modes and correspond to a 2×2 unitary matrix with determinant -1;[1] and *phaseshifters*, which act on one mode and simply apply a phase $e^{i\theta}$. Note that any unitary transformation acting on the one-photon Hilbert space automatically gets "lifted," by homomorphism, to a unitary transformation acting on the Hilbert space of n photons. Furthermore, every element of the *n-photon linear-optical group*—that is, every n-photon unitary transformation achievable using linear optics—arises in this way (see [11], Sec. III for details). Of course, if $n \geq 2$, then there are also n-photon unitaries that cannot be achieved linear-optically: that is, the n-photon linear-optical group is a proper subgroup of the full unitary group on the n-photon Hilbert space. In fact it is a very tiny subgroup of polynomial dimension, while the entire unitary group on this Hilbert space is of exponential dimension.

We call a set of optical gates S *universal* on m modes if it generates a dense subset of either $SU(m)$ (in the complex case) or $SO(m)$ (in the real case) when acting on a single photon. To clarify, if S is universal, this does *not* mean that linear optics with S is universal

[1] Some references use a different convention and assume that beamsplitters have determinant $+1$ [199]. Note that these two conventions are equivalent if one assumes that one can permute modes, i.e. apply the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ which has determinant -1.

for quantum computing! It only means that S densely generates the one-photon linear-optical group—or equivalently, the n-photon linear-optical group for any value of n. The latter kind of universality is certainly *relevant* for quantum computation: first, it already suffices for the boson sampling proposal of Aaronson and Arkhipov [11]; and second, if the single resource of adaptive measurements is added, then universal linear optics becomes enough for universal quantum computation, by the famous result of Knill, Laflamme, and Milburn (KLM) [172]. On the other hand, if we wanted to map a k-qubit Hilbert space *directly* onto an m-mode linear-optical Hilbert space, then as observed by Cerf, Adami and Kwiat [85], we would need $m \geq 2^k$ just for dimension-counting reasons.

Previously, Reck *et al.* [208] showed that the set of *all* phaseshifters and all beamsplitters is universal for linear optics, on any number of modes. Therefore it is natural to ask: is there *any* S set of beamsplitters and phaseshifters that generates a nontrivial set of linear-optical transformations, yet that still falls short of generating *all* of them? Here by "nontrivial," we simply mean that S does *something* more than permuting the modes around or adding phases to them.

If such a set S existed, we could then ask the *further* question of whether the n-photon subgroup generated by S was

(a) efficiently simulable using a classical computer, despite being nontrivial (much like the Clifford group for qubits),

(b) already sufficient for applications such as boson sampling and KLM, despite not being the full n-photon linear-optical group, or

(c) of "intermediate" status, neither sufficient for boson sampling and KLM nor efficiently simulable classically.

The implications for our dichotomy conjecture would of course depend on the answer to that further question.

In this chapter, however, we show that the further question never even arises, since *no such set S exists.* Indeed, any beamsplitter that acts nontrivially on two modes is universal on three or more modes. What makes this result surprising is that it holds *even if the beamsplitter angles are all rational multiples of π.* A priori, one might guess that by restricting the beamsplitter angles to (say) $\pi/4$, one could produce a linear-optical analogue of the Clifford group; but our result shows that one cannot.

Our proof uses standard representation theory and the classification of closed subgroups of $SU(3)$ [109, 143, 139]. From an experimental perspective, our result shows that any complex nontrivial beamsplitter suffices to create any desired optical network. From a computational complexity perspective, it implies a dichotomy theorem for optical gate sets: any set of beamsplitters or phaseshifters generates a set of operations that is either *trivially* classically simulable (even on n-photon input states), or else universal for quantum linear optics. In particular, any nontrivial beamsplitter can be used to perform boson sampling; there is no way to define an "intermediate" model of boson sampling[2] by restricting the allowed beamsplitters and phaseshifters.

Note that our result holds only for beamsplitters, i.e., optical gates that act on two modes and have determinant -1. We leave as an open problem whether our result can be extended to arbitrary two-mode gates, or to gates that act on three or more modes.

[2]Here by "intermediate," we mean computationally intermediate between classical computation and universal boson sampling.

Our work is the first that we know of to explore limiting the power of quantum linear optics by limiting the gate set. (However, after this work was published, Sawicki [213] reproved parts of our result using completely different techniques.) Previous work has considered varying the available input states and measurements. For example, as mentioned earlier, Knill, Laflamme, and Milburn [172] showed that linear optics with adaptive measurements is universal for quantum computation. Restricting to nonadaptive measurements seems to reduce the computational power of linear optics, but Aaronson and Arkhipov [11] gave evidence that the resulting model is still impossible to simulate efficiently using a classical computer. If Gaussian states are used as inputs and measurements are taken in the Gaussian basis only, then the model is efficiently simulable classically [51]; but with Gaussian-state inputs and *photon-number* measurements, there is recent evidence for computational hardness.[3] On the other hand, Oszmaniec and Zimborás recently considered the problem of adding additional operations to the set of linear optical transformations, and classified when they generate the entire unitary group on the much larger Hilbert space of n photons [202]. This result is complementary to ours as it addresses which operations boost you to full BQP universality, rather than our work which focuses on what operations can achieve universality in the linear optics sense.

One can view this work as a first as a first step toward proving the dichotomy conjecture (Conjecture 2.2.1) for *qubit*-based quantum circuits (i.e., the conjecture that every set of gates is either universal for quantum computation, or else universal for quantum computing under postselection, or else efficiently classically simulable). The tensor product structure of qubits gives rise to a much more complicated problem than the direct sum structure of linear optics. For that reason, one might expect the linear-optical "model case" to be easier to tackle first, and the results of this chapter confirms that expectation.

3.2 Background and Our Results

In a linear optical system with m modes, the state of a photon is described by a vector $|\psi\rangle$ in an m-dimensional Hilbert space. The basis states of the system are represented by strings $|s_1, s_2 \dots s_m\rangle$ where $s_i \in \{0, 1\}$ denotes the number of photons in the i^{th} mode, and $\Sigma_{j=1}^{m} s_j$ is the total number of photons (in this case, one). For example a one-photon, three-mode system has basis states $|100\rangle, |010\rangle$ and $|001\rangle$.

A k-local gate g is a $k \times k$ unitary matrix which acts on k modes at a time while acting in direct sum with the identity on the remaining $m - k$ modes. A *beamsplitter* b is a two-local gate with determinant -1. Therefore any beamsplitter has the form $b = \begin{pmatrix} \alpha & \beta^* \\ \beta & -\alpha^* \end{pmatrix}$ where $|\alpha|^2 + |\beta|^2 = 1$. Let b_{ij} denote the matrix action of applying the beamsplitter to modes i and j of a one-photon system. For example, if $m = 3$, we have that

$$b_{12} = \begin{pmatrix} \alpha & \beta^* & 0 \\ \beta & -\alpha^* & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad b_{31} = \begin{pmatrix} -\alpha^* & 0 & \beta \\ 0 & 1 & 0 \\ \beta^* & 0 & \alpha \end{pmatrix}$$

when written in the computational basis. A beamsplitter is called *nontrivial* if $|\alpha| \neq 0$ and $|\beta| \neq 0$, i.e. if the beamsplitter mixes modes.

We say that a set S of optical gates *densely generates* a continuous group G of unitary transformations, if the group H generated by S is a dense subgroup of G (that is, if $H \leq G$

[3] See http://www.scottaaronson.com/blog/?p=1579

and H contains arbitrarily close approximations to every element of G). Then we call S *universal on m modes* if it densely generates $SU(m)$ or $SO(m)$ when acting on m modes. (Due to the irrelevance of global phases, this is physically equivalent to generating $U(m)$ or $O(m)$ respectively.) In this definition we are assuming that whenever we have a k-mode gate in S, we can apply it to any subset of k modes (in any order), as often as desired. Note that we consider real $SO(m)$ evolutions to be universal as well; this is because the distinction between real and complex optical networks is mostly irrelevant[4] to computational applications of linear optics, such as the KLM protocol [172] and boson sampling [11].

A basic result in quantum optics, proved by Reck *et al.* [208], says that the collection of all beamsplitters and phaseshifters is universal. Specifically, given any target unitary U on m modes, there exists a sequence of $O(m^2)$ beamsplitters and phaseshifters whose product is exactly U. Reck *et al.*'s proof also shows an analogous result for real beamsplitters - namely, that any orthogonal matrix O can be written as the product of $O(m^2)$ real beamsplitters. Furthermore, it can easily be shown that there exist two beamsplitters b, b' whose products densely generate $O(2)$. Therefore b and b' can be used to simulate any real beamsplitter, and hence by Reck *et al.* [208], the set $\{b, b'\}$ is universal for linear optics.

In this chapter, we consider the universality of a *single* beamsplitter b. If b is trivial, then on m modes the matrices b_{ij} generates a subgroup of P_m, the set of $m \times m$ unitary matrices with all entries having norm zero or one. This is obviously non-universal, and the state evolutions on any number of photons are trivial to simulate classically. Our main result is that any nontrivial beamsplitter densely generates either all orthogonal transformations on three modes (in the real case), or all unitary transformations on three modes (in the complex case). From this, it follows easily from Reck *et al.* [208] that such a beamsplitter is also universal on m modes for any $m \geq 3$.

Theorem 3.2.1. *Let b be any nontrivial beamsplitter. Then the set $S = \{b_{12}, b_{13}, b_{23}\}$, obtained by applying b to all possible pairs among three photon modes,[5] densely generates either $SO(3)$ (if all entries of b are real) or $SU(3)$ (if any entry of b is non-real).*

Corollary 3.2.2. *Any nontrivial beamsplitter is universal on $m \geq 3$ modes.*

Proof. By Theorem 3.2.1, the set $S = \{b_{12}, b_{13}, b_{23}\}$ densely generates all orthogonal matrices with determinant 1. But since b has determinant -1, we know that S must generate all orthogonal matrices with determinant -1 as well.[6] Therefore, S densely generates the action of any real beamsplitter b' acting on two out of three modes. So by Reck et al [208], S also densely generates all orthogonal matrices on m modes for $m \geq 3$. □

Note that, although our proof of universality on three modes is nonconstructive, by the Solovay-Kitaev Theorem [99], there is an efficient algorithm that, given any target unitary U, finds a sequence of b's and their inverses approximating U up to error ε in $O\left(\log^{3.97}(\frac{1}{\varepsilon})\right)$

[4]The one case we know about where the real vs. complex distinction might matter is when using error-correcting codes. There, applying all possible orthogonal transformations to the physical modes or qubits might not suffice to apply all orthogonal transformations to the encoded modes or qubits. This could conceivably be an issue, for example, in the scheme of Gottesman, Kitaev, and Preskill [135] for universal quantum computing with linear optics.

[5]Technically, we could also consider the unitaries b_{21}, b_{31}, b_{32}, obtained by applying b to the same pairs of modes but reversing their order. However, this turns out not to give us any advantage.

[6]Indeed any orthogonal O with determinant -1 can be written as $O = b_{12}^{-1}O' = b_{12}O'$ for some O' of determinant 1.

time. Thus, our universality result also implies an efficient algorithm to construct any target unitary using beamsplitters and their inverses in the same manner as Reck *et al.* [208]. Wth polynomial overhead, one could achieve inverse exponential error in the target unitary - which is negligible for essentially all applications such as boson sampling or the KLM protocol.

We note that, as the Solovay-Kitaev algorithm requires the ability to apply the inverses of the beamsplitters as well. As mentioned previously it is a major open question to remove this limitation. If one has access to a single nontrivial beamsplitter b but not its inverse, then clearly with polynomial overhead, one could compile arbitrary optical transformations to inverse polynomial accuracy, simply by a volume argument. But if inverse polynomial accuracy is insufficient for the application, this could conceivably change the power of linear optics. We leave this as an open problem.

We now proceed to a proof of Theorem 3.2.1.

3.3 Proof of Main Theorem

We first consider applying a fixed beamsplitter

$$b = \begin{pmatrix} \alpha & \beta^* \\ \beta & -\alpha^* \end{pmatrix},$$

where α and β are complex and $|\alpha|^2+|\beta|^2 = 1$, to two modes of a three-mode optical system. We take pairwise products of these beamsplitter actions to generate three special unitary matrices. These three unitaries densely generate some group of matrices $G \leq SU(3)$. We then use the representation theory of subgroups of $SU(3)$ described in the work of Fairbairn, Fulton & Klink [109], Hanany & He [143], and Grimus & Ludl [139] to show that the beamsplitter must generate either all $SO(3)$ matrices (if the beamsplitter is real) or all $SU(3)$ matrices (if the beamsplitter has a complex entry).

Consider applying our beamsplitter to a three-mode system. Let R_1, R_2, R_3 be defined as the pairwise products of the beamsplitter actions below:

$$R_1 = b_{12}b_{13} = \begin{pmatrix} \alpha^2 & \beta^* & \alpha\beta^* \\ \alpha\beta & -\alpha^* & |\beta|^2 \\ \beta & 0 & -\alpha^* \end{pmatrix}, \quad R_2 = b_{23}b_{13} = \begin{pmatrix} \alpha & 0 & \beta^* \\ |\beta|^2 & \alpha & -\alpha^*\beta^* \\ -\alpha^*\beta & \beta & \alpha^{*2} \end{pmatrix},$$

$$R_3 = b_{12}b_{23} = \begin{pmatrix} \alpha & \alpha\beta^* & \beta^{*2} \\ \beta & -|\alpha|^2 & -\alpha^*\beta^* \\ 0 & \beta & -\alpha^* \end{pmatrix}.$$

Since R_1, R_2, R_3 are even products of matrices of determinant -1, they are all elements of $SU(3)$. Let $G \leq SU(3)$ be the subgroup densely generated by products of the elements $\{R_1, R_2, R_3\}$ and their inverses[7]. Let G_M be the set of matrices representing G under this construction. First we will show that these matrices G_M form an irreducible representation of G.

Claim 1. *The set $\{R_1, R_2, R_3\}$ generates an irreducible three-dimensional representation of G.*

[7]Since $b_{ij}^{-1} = b_{ij}$, the beamsplitter is capable of generating the inverses of the R_i as well.

Proof. Suppose that some matrix

$$U = \begin{pmatrix} A & D & G \\ B & E & H \\ C & F & I \end{pmatrix}$$

commutes with R_1, R_2, and R_3. Then we claim that U is a constant multiple of the identity, i.e. $A = E = I$ and $D = G = H = B = C = F = 0$.

From the claim, it follows easily that the representation is irreducible. Indeed, suppose the representation is reducible, so preserves a non-trivial subspace. Since our representation is unitary, this implies that our representation is decomposable, i.e. by a change of basis it can be brought into block-diagonal form.[8] In the new basis, the matrix consisting of 1's on the diagonal in the first block, and 2's in the diagonal of the second block, commutes with all elements of G, and in particular with R_1, R_2, R_3. But that matrix is not a multiple of the identity. Hence if only multiples of the identity commute with R_1, R_2, R_3, the representation must be irreducible.

We now prove the claim. First, since U commutes with R_1,

$$\begin{pmatrix} A & D & G \\ B & E & H \\ C & F & I \end{pmatrix} \begin{pmatrix} \alpha^2 & \beta^* & \alpha\beta^* \\ \alpha\beta & -\alpha^* & |\beta|^2 \\ \beta & 0 & -\alpha^* \end{pmatrix} = \begin{pmatrix} \alpha^2 & \beta^* & \alpha\beta^* \\ \alpha\beta & -\alpha^* & |\beta|^2 \\ \beta & 0 & -\alpha^* \end{pmatrix} \begin{pmatrix} A & D & G \\ B & E & H \\ C & F & I \end{pmatrix}.$$

This imposes nine equations. Below we give the equations coming from the (1,1), (1,2), (2,2), (2,3), and (3,2) entries of the above matrices respectively.

$$(D\alpha + G)\beta = (C\alpha + B)\beta^*, \tag{3.1}$$

$$(A - E - F\alpha)\beta^* = D(\alpha^2 + \alpha^*), \tag{3.2}$$

$$B\beta^* = D\alpha\beta + F\beta\beta^*, \tag{3.3}$$

$$B\alpha\beta^* + E\beta\beta^* - H\alpha^* = G\alpha\beta - H\alpha^* + I\beta\beta^*, \tag{3.4}$$

$$C\beta^* = D\beta. \tag{3.5}$$

Note that Eqs. (3.5) and (3.1) imply that

$$G\beta = B\beta^*. \tag{3.6}$$

So by Eq. (3.4) we have

$$E\beta\beta^* = I\beta\beta^*. \tag{3.7}$$

So since $0 < |\beta| < 1$, we have $I = E$.

In total so far we have $I = E$, $G\beta = B\beta^*$ and $C\beta^* = D\beta$.

Next, since U commutes with R_2,

$$\begin{pmatrix} A & D & G \\ B & E & H \\ C & F & E \end{pmatrix} \begin{pmatrix} \alpha & 0 & \beta^* \\ |\beta|^2 & \alpha & -\alpha^*\beta^* \\ -\alpha^*\beta & \beta & \alpha^{*2} \end{pmatrix} = \begin{pmatrix} \alpha & 0 & \beta^* \\ |\beta|^2 & \alpha & -\alpha^*\beta^* \\ -\alpha^*\beta & \beta & \alpha^{*2} \end{pmatrix} \begin{pmatrix} A & D & G \\ B & E & H \\ C & F & E \end{pmatrix}.$$

[8] To see the equivalence of "reducible" and "decomposable" for unitary representations, it suffices to note that, if a set of unitary matrices always map a subspace V to itself, then they cannot map any vector not in V to a vector in V, since this would violate unitarity.

This imposes another nine equations. Here are the equations from the (1,1), (2,1) and (2,2) entries respectively, which we have simplified using $I = E$, $G\beta = B\beta^*$ and $C\beta^* = D\beta$:

$$D\beta = D\beta\beta^* - G\alpha^*\beta, \tag{3.8}$$

$$E\beta\beta^* - H\alpha^*\beta = A\beta\beta^* - C\alpha^*\beta^*, \tag{3.9}$$

$$H\beta = D\beta\beta^* - F\alpha^*\beta^*. \tag{3.10}$$

Note that Eqs. (3.8) and (3.10), combined with the fact that $G\beta = B\beta^*$, imply that $D\beta = H\beta$, and hence $D = H$.

Plugging this in to Eq. (3.9), we see that $E\beta\beta^* - D\alpha^*\beta = A\beta\beta^* - C\alpha^*\beta^*$. Using $C\beta^* = D\beta$ these last two terms cancel, so $E\beta\beta^* = A\beta\beta^*$, and hence $E = A$. So overall we have established that $A = E = I$, $D = H$, $B = F$, $G\beta = B\beta^*$ and $C\beta^* = D\beta$.

Now suppose $B = 0$. Then we have from above that $B = F = G = 0$. By Eq. (3.8) we also have $D\beta = D\beta\beta^* \Rightarrow D = 0$ since $0 < |\beta| < 1$. Hence we have $C = 0$ as well by the fact that $C\beta^* = D\beta$. Therefore U is a multiple of the identity, as desired.

So it suffices to prove that $B = 0$. Suppose $B \neq 0$; then we will derive a contradiction. Since U commutes with R_3,

$$\begin{pmatrix} A & D & G \\ B & A & D \\ C & B & A \end{pmatrix} \begin{pmatrix} \alpha & \alpha\beta^* & \beta^{*2} \\ \beta & -|\alpha|^2 & -\alpha^*\beta^* \\ 0 & \beta & -\alpha^* \end{pmatrix} = \begin{pmatrix} \alpha & \alpha\beta^* & \beta^{*2} \\ \beta & -|\alpha|^2 & -\alpha^*\beta^* \\ 0 & \beta & -\alpha^* \end{pmatrix} \begin{pmatrix} A & D & G \\ B & A & D \\ C & B & A \end{pmatrix}.$$

This imposes yet another nine equations, but we will only need the one coming from the (2,2) entry of the above matrices to complete the proof:

$$B\alpha\beta^* = -B\alpha^*\beta^*. \tag{3.11}$$

Since $B \neq 0$, Eq. (3.11) implies that $\alpha = -\alpha^*$, i.e. α is pure imaginary. Furthermore, since $G\beta = B\beta^*$, we have $G \neq 0$ as well. Using this, we can write out Eqs. (3.2) and (3.3) as follows:

$$(-B\alpha)\beta^* = D(\alpha^2 - \alpha) \quad \Rightarrow \quad G\beta = D(1 - \alpha), \tag{3.12}$$

$$B\beta^* = D\alpha\beta + F\beta\beta^* \quad \Rightarrow \quad G = D\alpha + G\beta. \tag{3.13}$$

Summing these equations, we see that $G = D$. Plugging back into Eq. (3.13), we see that $\beta = 1 - \alpha$. Since α is pure imaginary this contradicts $|\alpha|^2 + |\beta|^2 = 1$.

To summarize, if U commutes with all elements of G, then U is a multiple of the identity. This proves the claim and hence the theorem. □

We have learned that the set G_M forms a three-dimensional irreducible representation of G. We now leverage this fact, along with the classification of finite subgroups of $SU(3)$, to show that G is not finite.

Claim 2. *G is infinite.*

Proof. By Claim 1, if G is finite then $\{R_1, R_2, R_3\}$ generates an irreducible representation of G. The finite subgroups of $SU(3)$ consist of the finite subgroups of $SU(2)$, 12 exceptional finite subgroups, and two infinite families of "dihedral-like" groups, whose irreducible representations are classified in [109, 143, 139]. Our proof proceeds by simply enumerating

the possible finite groups that G could be, and showing that $\{R_1, R_2, R_3\}$ cannot generate an irreducible representation of any of them.

First we eliminate the possibility that G is an exceptional finite subgroup of $SU(3)$. Of the 12 exceptional subgroups, only eight of them have three-dimensional irreps: they are labeled $\Sigma(60)$, $\Sigma(60) \times \mathbb{Z}_3$, $\Sigma(168)$, $\Sigma(168) \times \mathbb{Z}_3$, $\Sigma(216)$, $\Sigma(36 \times 3)$, $\Sigma(216 \times 3)$, and $\Sigma(360 \times 3)$. So by Claim 1, if G is finite and exceptional, then it is one of these eight groups.

The character tables of these groups are provided in [109] and [143]. Recall that the character of an element of a representation is the trace of its representative matrix. The traces of the matrices R_1, R_2, R_3, denoted T_1, T_2, T_3, are given by

$$T_1 = \alpha^2 - 2\alpha^* \tag{3.14}$$

$$T_2 = (\alpha^*)^2 + 2\alpha \tag{3.15}$$

$$T_3 = -|\alpha|^2 + \alpha - \alpha^* = -|\alpha|^2 + 2\mathrm{Im}(\alpha) \tag{3.16}$$

We will show that these cannot be the characters of the elements of a three-dimensional irrep of $\Sigma(60)$, $\Sigma(60) \times \mathbb{Z}_3$, $\Sigma(168)$, $\Sigma(168) \times \mathbb{Z}_3$, $\Sigma(216)$, $\Sigma(36 \times 3)$, $\Sigma(216 \times 3)$, or $\Sigma(360 \times 3)$.

There are two three-dimensional irreps of $\Sigma(60)$ up to conjugation [109]. The characters of their elements all lie in the set $\left\{0, -1, 3, \frac{1+\sqrt{5}}{2}, \frac{1-\sqrt{5}}{2}\right\}$. Note that $0 < |\alpha|^2 < 1$, which means that T_3 cannot be in this set unless $T_3 = \frac{1-\sqrt{5}}{2}$ and $\mathrm{Im}(\alpha) = 0$. But then this implies $\alpha = \pm\sqrt{\frac{\sqrt{5}-1}{2}}$. Plugging this into T_1 and T_2, we see they are not in the set of allowed values. Hence G is not $\Sigma(60)$.

The characters of the three-dimensional irreps of $\Sigma(60) \times \mathbb{Z}_3$ are identical to those of $\Sigma(60)$, but with the additional possibility that they can be multiplied by $e^{\frac{2\pi i}{3}}$ or $e^{\frac{4\pi i}{3}}$. The same argument as above shows that in order for T_3 to be in the set of allowed characters, we must have $\alpha = \pm\sqrt{\frac{\sqrt{5}-1}{2}}$ or $\alpha = \frac{1+\sqrt{5}}{8}\left(\pm 1 \pm \sqrt{3}i\right)$. Plugging these into T_1, we see the possible values of T_1 are not in the set of allowed characters, hence G is not $\Sigma(60) \times \mathbb{Z}_3$.

There are two three-dimensional irreps of $\Sigma(168)$ up to conjugation [109]. The characters of their elements all lie in the set $S = \left\{0, \pm 1, 3, \frac{1}{2}(-1 \pm i\sqrt{7})\right\}$. Since $0 < |\alpha|^2 < 1$, if T_3 is in this set it must have value $\frac{1}{2}(-1 \pm i\sqrt{7})$. Therefore we must have $\alpha = \pm\frac{1}{4} \pm \frac{\sqrt{7}}{4}i$. This implies that $\alpha^2 = \frac{-3}{8} \pm \frac{\sqrt{7}}{8}i$ and $2\alpha^* = \pm\frac{1}{2} \pm \frac{\sqrt{7}}{2}i$. Regardless of the signs chosen, this means that T_1 is not in the set S of allowed values. Hence G is not $\Sigma(168)$.

The characters of the three-dimensional irreps of $\Sigma(168) \times \mathbb{Z}_3$ are identical to those of $\Sigma(168)$, but with the additional possibility that they can be multiplied by $e^{\frac{2\pi i}{3}}$ or $e^{\frac{4\pi i}{3}}$. The same argument as above shows that in order for T_3 to be in the set of allowed characters, we must have $\alpha = \pm\frac{1}{4} \pm \frac{\sqrt{7}}{4}i$, $\alpha = \frac{1}{4}(\pm\sqrt{5} \pm i\sqrt{3})$, or $\alpha = \pm\frac{1}{4\sqrt{2}}\sqrt{7\sqrt{21} - 13} \pm \frac{\sqrt{3}+\sqrt{7}}{8}i$. Plugging these into T_1, we see that the possible values of T_1 are not in the set of allowed characters, hence G is not $\Sigma(168) \times \mathbb{Z}_3$.

There is one three-dimensional irrep of $\Sigma(216)$ up to conjugation [109]. The characters

of its elements all lie in the set $\{0, -1, 3\}$. Since T_3 cannot be in this set, G is not $\Sigma(216)$.

There are eight three-dimensional irreps of $\Sigma(36\times3)$ up to conjugation [143]. The characters of their elements all lie in the set $S = \{0, \pm1, \pm e_3, \pm e_3^2, \pm e_4, \pm e_{12}^7, \pm e_{12}^{11} \pm 3, \pm 3e_3, \pm 3e_3^2\}$ where $e_n = e^{\frac{2\pi i}{n}}$. Since $\mathrm{Re}(T_3) = -|\alpha|^2$ and $0 < |\alpha|^2 < 1$, if $T_3 \in S$ then we must have $T_3 \in \{\pm e_3, \pm e_3^2, \pm e_{12}^7, \pm e_{12}^{11}\}$. Solving for α gives us $\alpha \in \left\{ \frac{\pm\sqrt{5} \pm \sqrt{3}i}{4}, \frac{\pm\sqrt{8\sqrt{3}-1} \pm i}{4} \right\}$. A straightforward evaluation of possible values of T_1 shows $T_1 \notin S$. So T_1 and T_3 cannot be characters of these irreps, and hence G is not $\Sigma(36 \times 3)$.

There are seven three-dimensional irreps of $\Sigma(216 \times 3)$ up to conjugation [143]. The characters of their elements all lie in the set

$$S = \{0, \pm1, 3, \pm e_3, \pm e_3^2, -e_9^2, -e_9^4, -e_9^5, -e_9^7, \pm e_9^2 + e_9^5, 2e_9^2 + e_9^5, -e_9^2 - 2e_9^5, e_9^4 + e_9^7, e_9^4 + 2e_9^7, -2e_9^4 - e_9^7\}.$$

If $T_3 \in S$, then for each case we can solve for α and hence T_1. As above, a straightforward calculation shows that for no $T_3 \in S$ do we have $T_1 \in S$. Hence G is not $\Sigma(216 \times 3)$.

There are four three-dimensional irreps of $\Sigma(360 \times 3)$ up to conjugation [143]. The characters of their elements all lie in the set

$$S = \{0, \pm1, \pm e_3, \pm e_3^2, 3e_3, 3e_3^2, -e_5 - e_5^4, -e_5^2 - e_5^3, -e_{15} - e_{15}^4, -e_{15}^7 - e_{15}^{13}, -e_{15}^{11} - e_{15}^{14}, -e_{15}^2 - e_{15}^8\}.$$

Again a straightforward calculation shows that for no $T_3 \in S$ do we have $T_1 \in S$. Hence G is not $\Sigma(360 \times 3)$.

We have therefore shown that G_M is not an irrep of an exceptional finite subgroup of $SU(3)$.

Next we will show that G_M is not in one of the two infinite families of "dihedral-like" finite subgroups of $SU(3)$, known as the C-series and the D-series groups. The most well-known members of these series are $\Delta(3n^2)$ and $\Delta(6n^2)$, labeled by $n \in \mathbb{N}$, which consist of all 3 by 3 even permutation matrices (for $\Delta(3n^2)$) or all 3 by 3 permutation matrices (for $\Delta(6n^2)$) whose entries are replaced by nth roots of unity. In early works describing subgroups of $SU(3)$, such as Fairbairn, Fulton, and Klink [109], only $\Delta(3n^2)$ and $\Delta(6n^2)$ appear as elements of these series. However in 2011, Ludl [183] pointed out that there exist nontrivial subgroups of $\Delta(3n^2)$ and $\Delta(6n^2)$ which are missing from these references. Fortunately these groups have now been fully classified [138], and sufficient constraints have been placed on their representations [139] that we can eliminate the possibility that G is an irrep of any C or D-series group.

In the following, we first eliminate the possibility that G_M is an irrep of $\Delta(3n^2)$ or $\Delta(6n^2)$ following the work of Fairbairn, Fulton, and Klink [109]. Afterwards we show that these arguments suffice to prove G_M is not an irrep of any of the C-series or D-series groups, using the work of Grimus and Ludl [139].

The three-dimensional irreps of $\Delta(3n^2)$ are labeled by integers $m_1, m_2 \in \{0, \ldots, n-1\}$, and have conjugacy classes labeled by types A, C, E and numbers $p, q \in \{0, \ldots, n-1\}$. The respective characters are either 0 for conjugacy classes $C(p,q)$ and $E(p,q)$ or

$$e^{\frac{2\pi i}{n}(m_1 p + m_2 q)} + e^{\frac{2\pi i}{n}(m_1 q - m_2 (p+q))} + e^{\frac{2\pi i}{n}(-m_1 (p+q) + m_2 p)} \tag{3.17}$$

for conjugacy class $A(p,q)$.

Assume that G_M is an irrep of $\Delta(3n^2)$ for some n—we will derive a contradiction shortly. Then the trace of each R_i must be zero (if R_i is a representative of type C or E) or of the

form of Eq. (3.17) (if R_i is a representative of type A). However, we can show that none of the traces T_i can be 0 because our beamsplitter is nontrivial. Indeed T_3 cannot be zero as $0 < |\alpha|^2 < 1$. We know that in order for T_1 to be zero, we need $\alpha^2 = 2\alpha^*$, which implies $|\alpha| = 2$ which is not possible, and likewise with T_2. Hence each T_i must have the form of Eq. (3.17), which implies each R_i is in conjugacy class $A(p_i, q_i)$ for some choice of p_i, q_i. However, looking at the multiplication table for this group provided in [109, Table VIII], we have that $A(p,q)A(p',q') = A\left(p+q \mod n, p'+q' \mod n\right)$. Hence the T_i's cannot possibly generate all of $\Delta(3n^2)$ for any n, since they cannot generate elements in the conjugacy classes $C(p,q)$ or $E(p,q)$. This contradicts our assumption that the R_i's generate an irrep of $\Delta(3n^2)$. Therefore G_M is not an irrep of $\Delta(3n^2)$ for any n.

We now extend this argument to eliminate the possibility that G_M is an irrep of any of the C-series groups. In Appendix E of [139], Grimus and Ludl show that for any three-dimensional irrep of a C-series group, there exists a basis (and an ordering of that basis) in which all elements of the A conjugacy classes are represented by diagonal matrices and $E(0,0)$ is represented by

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{3.18}$$

From this it can be easily shown that in any three-dimensional irrep of a C-series group, all elements of types C and E are represented by traceless matrices.[9] In our previous arguments eliminating $\Delta(3n^2)$ as a possibility, we showed that none of the generators R_i can be traceless, so each of the R_i must be of type A. Again this is a contradiction since elements of type A generate an abelian group and G_M is nonabelian. Hence G_M cannot be an irrep of one of the C-series groups.

Next we turn our attention to the D-series finite subgroups of $SU(3)$. We begin by showing that G_M cannot be an irrep of $\Delta(6n^2)$ for any n, and we will later generalize this to eliminate all D-series groups as possibilities. The group $\Delta(6n^2)$ contains 6 families of conjugacy classes, labeled by types A, B, C, D, E, F and by integers p, q as above. The three-dimensional irreps of $\Delta(6n^2)$ are again labeled by (m_1, m_2), which now take values in $(m,0), (0,m)$ or (m,m), as well as $t \in \{0,1\}$. The character of each element is

$$\mathrm{Tr}(A(p,q)) = e^{\frac{2\pi i}{n}(m_1 p + m_2 q)} + e^{\frac{2\pi i}{n}(m_1 q - m_2(p+q))} + e^{\frac{2\pi i}{n}(-m_1(p+q) + m_2 p)} \tag{3.19}$$

$$\mathrm{Tr}(B(p,q)) = (-1)^t e^{\frac{2\pi i}{n}(m_1 p + m_2 q)} \tag{3.20}$$

$$\mathrm{Tr}(D(p,q)) = (-1)^t e^{\frac{2\pi i}{n}\left(m_1(\frac{n}{2} - p - q) + m_2 p\right)} \tag{3.21}$$

$$\mathrm{Tr}(F(p,q)) = (-1)^t e^{\frac{2\pi i}{n}\left(m_1 q + m_2(\frac{n}{2} - p - q)\right)} \tag{3.22}$$

$$\mathrm{Tr}(C(p,q)) = \mathrm{Tr}(E(p,q)) = 0 \tag{3.23}$$

We now eliminate the possibility that G_M is an irrep of $\Delta(6n^2)$ for any n. Again assume by way of contradiction that G_M is an irrep of $\Delta(6n^2)$ for some n. Then each R_i must be in one of the types A, B, C, D, E, F, and each trace T_i must have the corresponding character from Eqs. (3.19)-(3.23). As noted previously each T_i cannot be 0, so in fact each R_i must

[9] To see this, note that by the group multiplication table in [13] Table VIII, we have that $A(p,q) = E(p,q)E(0,0)E(0,0)$, so $A(p,q)$ is in a D-series group if and only if $E(p,q)$ is in the group. Additionally, since $A(p,q)E(0,0) = E(p,q)$, all elements of type E are obtainable by multiplying an element of type A by $E(0,0)$. Since in this basis the A matrices are diagonal, and $E(0,0)$ is represented by the above matrix (3.18), this implies the claim for elements of type E. A similar argument holds for the elements of type C.

be of type A, B, D or F. Furthermore, we will show the following Lemma:

Lemma 3.3.1. *If G_M is an irrep of $\Delta(6n^2)$, then all R_i of types B, D or F are of the same type.*

By Lemma 3.3.1, some of the R_i's belong to a single type B, D or F while the remaining R_i's are of type A. However, by examining the multiplication table for this group provided in [109, Table VIII], one can see that any number of elements of type A plus any number of elements from a single type B, D, or F cannot generate the entire group. This contradicts our assumption that the R_i's generate an irrep of $\Delta(6n^2)$. Hence G_M is not an irrep of $\Delta(6n^2)$ so G cannot be $\Delta(6n^2)$ by Claim 1.

We now prove Lemma 3.3.1 before continuing the proof of Claim 2.

Proof of Lemma 3.3.1. Assume that G_M is an irrep of $\Delta(6n^2)$. We will show that all of the R_i's of types B, D or F are of the same type. We proceed by enumerating all pairs R_i, R_j for $i \neq j$ and show that it is not possible for both R_i and R_j to be of distinct types B, D or F.

Let $\alpha = a + bi$ where a and b are real. If R_i is of type B, D or F, then T_i has norm 1, which imposes the following equations on a and b:

$$|T_1|^2 = 1 \Rightarrow (a^2+b^2)^2 + 4[a^2(1-a) + b^2(3+a)] = 1 \tag{3.24}$$

$$|T_2|^2 = 1 \Rightarrow (a^2+b^2)^2 + 4[a^2(1+a) + b^2(1-3a)] = 1 \tag{3.25}$$

$$|T_3|^2 = 1 \Rightarrow (a^2+b^2)^2 + 4b^2 = 1 \tag{3.26}$$

First suppose that R_1 and R_2 are members of distinct types B, D, or F. Then $|T_1| = |T_2| = 1$. The only solutions to Eqs. (3.24) and (3.25) in which $0 < |\alpha|^2 = a^2 + b^2 < 1$ are $\left(a = 0, b = \pm\sqrt{\sqrt{5}-2}\right)$ and $\left(a = \pm\frac{1}{2}\sqrt{3(\sqrt{5}-2)}, b = \pm\frac{1}{2}\sqrt{\sqrt{5}-2}\right)$. Note also that the product R_1R_2 must be of type C or E according to the group multiplication table in [109, Table VIII]. Hence the trace of R_1R_2 must be 0 if G_M is an irrep of $\Delta(6n^2)$. This implies that

$$\mathrm{Tr}(R_1R_2) = \alpha^3 - \alpha^{*3} + |\beta|^2\left(1 + \beta + \beta^* - |\alpha|^2\right) - |\alpha|^2 = 0 \tag{3.27}$$

Since we have $\alpha = a + bi$ where the values of a and b are one of the six possibilities above, one can see that there is no β which satisfies Eq. (3.27). Indeed, note that $\alpha^3 - \alpha^{*3}$ is nonzero and pure imaginary, while the rest of the expression is real, so the terms in Eq. (3.27) cannot sum to zero. This provides the desired contradiction. We conclude that R_1 and R_2 cannot be of distinct types B, D, or F.

Next suppose that R_1 and R_3 are of distinct types B, D or F. Then $|T_1| = |T_3| = 1$. If $\alpha = a + bi$ as before, Eqs. (3.24) and (3.26), combined with the fact that $0 < |\alpha|^2 = a^2 + b^2 < 1$, imply that $a = 0$ and $b = \pm\sqrt{\sqrt{5}-2}$. Again, using the group multiplication table in [109, Table VIII] we must have that R_1R_3 is of type C or E so

$$\mathrm{Tr}(R_1R_3) = \alpha^3 + \alpha^*|\alpha|^2 + \alpha^{*2} + |\beta|^2(1 + \beta + \beta^* + \alpha^2) = 0 \tag{3.28}$$

Since $\alpha = \pm i\sqrt{\sqrt{5}-2}$, this is a contradiction—for the terms $\alpha^3 + \alpha^*|\alpha|^2$ of Eq. (3.28) are nonzero and pure imaginary while the remaining terms are real. Hence R_1 and R_3 cannot be of distinct types B, D, or F.

Finally suppose that R_2 and R_3 are of distinct types B, D, or F. Then $|T_2| = |T_3| = 1$. If $\alpha = a+bi$ then the only solutions to Eqs. (3.25) and (3.26) in which $0 < |\alpha|^2 = a^2+b^2 < 1$ are $\left(a = 0, b = \pm\sqrt{\sqrt{5}-2}\right)$ and $(a \approx 0.437668, b \approx \pm 0.457975)$. Furthermore using the group multiplication table in [109, Table VIII] we must have that R_2R_3 is of type C or E so

$$\mathrm{Tr}(R_2R_3) = \alpha^2 - \alpha^{*3} - \alpha|\alpha|^2 + |\beta|^2(\alpha\beta^* - \alpha^*\beta^* - 2\alpha^*) = 0 \tag{3.29}$$

With slightly more work, one can again check that Eq. (3.29) cannot be satisfied with the above values of α, under the additional constraint that $|\alpha|^2+|\beta|^2 = 1$, providing the desired contradiction. Hence R_2 and R_3 cannot be of distinct types B, D, or F, which completes the proof of Lemma 3.3.1.

□

We have therefore eliminated the possibility that G_M is an irrep of $\Delta(6n^2)$ for any n, and so $G \neq \Delta(6n^2)$ by Claim 1.

We now extend this argument to eliminate the possibility that G_M is an irrep of any of the D-series groups. In Appendix G of [139], Grimus and Ludl show that for any three-dimensional irrep of a D-series group, there exists a basis (and an ordering of that basis) in which all elements of the A conjugacy classes are represented by diagonal matrices, and $E(0,0)$ and $B(0,0)$ are represented by

$$E(0,0) \to \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad B(0,0) \to \pm \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{3.30}$$

From this it can be easily shown that in any three-dimensional irrep of a D-series group, all elements of types C and E are represented by traceless matrices, and all elements of types B, D and F are represented by matrices whose trace has unit norm.[10] In our previous arguments eliminating $\Delta(6n^2)$ as a possibility, we showed none of our generators R_i can be traceless, and those of trace norm 1 are of the same type. Hence some of our generators are of type A and the remainder are of a single type B, D or F. Again this is a contradiction since any number of elements of type A and any number of elements of a single type B, D or F do not suffice to generate any D-series group - in particular they cannot generate $E(0,0)$. Hence G_M cannot be an irrep of any of the D-series groups. This concludes the proof that G_M cannot be an irrep of any of the "dihedral-like" subgroups of $SU(3)$.

Finally we will show that G is not a finite subgroup of $SU(2)$. Since $SU(2)$ is a double cover of $SO(3)$, if G is a finite subgroup of $SU(2)$, then G must be either a finite subgroup of $SO(3)$ or else the double cover of such a subgroup. We first eliminate the finite subgroups of $SO(3)$. The dihedral and cyclic subgroups have no three-dimensional irreps; hence G cannot be one of these by Claim 1. The icosahedral subgroup is isomorphic to $\Sigma(60)$ so has already been eliminated. The octahedral and tetrahedral subgroups do have three-dimensional irreps. However, the characters of their elements all lie in the set $\{0, \pm 1, \pm 3\}$, so these can be eliminated just as the exceptional groups of $SU(3)$ were eliminated.

[10] The fact that matrices representing elements of types C and E are traceless follows from the previous arguments regarding C-series groups. The fact that matrices representing elements of types B, D and F have traces of norm 1 follows by an identical argument since $A(p,q) = B(p,q)B(0,0)$, $A(p,q)B(0,0) = B(p,q)$, A matrices have diagonal representatives, and $B(0,0)$ is represented by the above in this basis. A similar argument holds for elements of types D and F.

Now all that remains are double covers of the finite subgroups of $SO(3)$. The binary dihedral groups, also known as the dicyclic groups, have no three-dimensional irreps, so G cannot be a binary dihedral group by Claim 1. The binary tetrahedral group has one three-dimensional irrep, with character values in the set $\{0, \pm 1, \pm 3\}$. So T_3 cannot be in this set as noted above.

The binary octahedral group has two three-dimensional irreps, with character values also in $\{0, \pm 1, \pm 3\}$, so is likewise eliminated. The binary icosahedral group has two three-dimensional irreps, with all characters in the set $\{0, -1, 3, \frac{\sqrt{5} \pm 1}{2}\}$. As discussed in the case of $\Sigma(60)$, our traces cannot take these values.

In summary, by enumeration of the finite subgroups of $SU(3)$, we have shown that G cannot be finite. □

Corollary 3.3.2. *G is a continuous (Lie) subgroup of $SU(3)$.*

Proof. G is infinite by Claim 2. Furthermore G is closed because it is the set of matrices *densely* generated by $\{R_1, R_2, R_3\}$. It is well-known that a closed, infinite subgroup of a Lie group is also a Lie group (this is Cartan's theorem [83]). The corollary follows. □

Next we show that G must be either $SO(3)$, $SU(2)$ or $SU(3)$. Furthermore, the set of matrices G_M densely generated by $\{R_1, R_2, R_3\}$ consists of either all $SO(3)$ matrices or all $SU(3)$ matrices.

Claim 3. *G is either $SO(3)$, $SU(2)$, or $SU(3)$. Furthermore, G_M consists of either all 3×3 special unitary matrices (if the beamsplitter b has a non-real entry), or all 3×3 special orthogonal matrices (if b is real).*

Proof. Since R_1, R_2, and R_3 do not commute, G is nonabelian. By Corollary 3.3.2, we know G is a Lie group, and furthermore G is closed. The nonabelian closed connected Lie subgroups of $SU(3)$ are well-known [76]: they are $SU(3)$, $SU(2) \times U(1)$, $SU(2)$, and $SO(3)$. Meanwhile, the closed disconnected Lie subgroups of $SU(3)$ are $\Delta(3\infty)$ and $\Delta(6\infty)$, as described in [109].

Note that $\Delta(3\infty)$ and $\Delta(6\infty)$ are the analogues of $\Delta(3n^2)$ and $\Delta(6n^2)$ as $n \to \infty$. Our above arguments showing that $G \neq \Delta(3n^2)$ and $G \neq \Delta(6n^2)$ carry over in this limit, because at no point did we use the fact that n or m were finite. Therefore G cannot be either of these continuous groups.

By Claim 1, G has a three-dimensional irrep. Of the remaining groups, only $SU(2)$, $SO(3)$, and $SU(3)$ have three-dimensional irreps. Furthermore, it is well known that the only three-dimensional irrep of $SU(2)$ is as $SO(3)$. This is because $SU(2)$ has exactly one irrep in each finite dimension (See [76, Section II.5] or [245] for details), and $SU(2)$ has an obvious representation as $SO(3)$ via the fact that $SU(2)$ is a double cover of $SO(3)$. Since we are only concerned with the set of matrices G_M generated, without loss of generality we can assume G is either $SO(3)$ or $SU(3)$.

It is well-known that the only three-dimensional irrep of $SU(3)$ is the natural one, as the group of all 3×3 special unitary matrices ([76, Section VI.5]). Likewise, the only three-dimensional irrep of $SO(3)$ is the natural one, up to conjugation by a unitary [76]. Hence G_M consists of either all 3×3 special unitary matrices (case A), or all 3×3 special orthogonal matrices conjugated by some unitary U (case B).

We now show that if the beamsplitter b is real, then we are in case B and without loss of generality the conjugating unitary U is real. Hence G_M is the set of all 3×3 orthogonal

matrices. Otherwise, if b has a complex entry, we will show we are in case A and G_M is the set of all 3×3 special unitary matrices.

First, suppose b is real. Then all matrices in our generating set are orthogonal, so all matrices in G_M are orthogonal. Hence we are in case B, and since all matrices in G_M are real, without loss of generality U is a real matrix as well.

Now suppose that b has a complex entry. Then either α or β are not real. First, suppose α is not real. Then $\mathrm{Tr}(R_1) = \alpha^2 - 2\alpha^*$ is not real because $0 < |\alpha| < 1$. But since conjugating a matrix by a unitary preserves its trace, and we are in case B, the traces of all matrices in G_M must be real. In particular $\mathrm{Tr}(R_1)$ must be real, which is a contradiction. Therefore if α is not real then we must be in case A.

Next, suppose β is not real. Then we can obtain a similar contradiction. Let $\beta = p + qi$ where p and q are real. By direct calculation one can show that $\mathrm{Im}\left(\mathrm{Tr}(R_1 R_2 R_3 R_1)\right) = |\beta|^4 \left(\beta^{*2} + 2\beta\right)$. Since our beamsplitter is nontrivial, $|\beta|^4 \neq 0$, so this quantity is 0 if and only if $\beta^{*2} + 2\beta = 0 \Leftrightarrow 2q(1-p) = 0$. But this cannot occur, since $q \neq 0$ (because β is not real), and $1 - p \neq 0$ (because the beamsplitter is nontrivial). Hence in this case $\mathrm{Tr}(R_1 R_2 R_3 R_1)$ is imaginary, which contradicts the fact we are in case B. Therefore if β is not real then we must be in case A, which completes the proof. □

Theorem 3.2.1 follows from Claim 3. Having proved our main result, we can now easily show two alternative versions of the theorem as well.

Corollary 3.3.3. *Any nontrivial two-mode optical gate $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ (not necessarily of determinant -1), plus the set of all phaseshifters densely generates $SU(m)$ on $m \geq 3$ modes.*

Proof. Since g is unitary we have $\det(g) = e^{i\theta}$ for some θ. By composing g with a phase of $e^{i\frac{\pi-\theta}{2}}$, we obtain a nontrivial beamsplitter g' of determinant -1. The gate g' is universal by Theorem 3.2.1, hence this gate set is universal as well. □

Corollary 3.3.4. *Any nontrivial two-mode real optical gate g is universal for quantum linear optics.*

Proof. Since g is real, g must have determinant ± 1. The case of $\det(g) = -1$ is handled by Theorem 3.2.1, so we now prove the $\det(g) = +1$ case. In this case g is a rotation by an angle θ. The fact that g is nontrivial means θ is not a multiple of $\pi/2$. The beamsplitter actions b_{12}, b_{23}, b_{13} can be viewed as three-dimensional rotations by angle θ about the x, y and z axes. So the question reduces to "For which angles θ (other than multiples of $\pi/2$) do rotations by θ about the x, y and z axes fail to densely generate all possible rotations?"

This question is easily answered using the well-known classification of closed subgroups of $SO(3)$. The finite subgroups of $SO(3)$ are the cyclic, dihedral, tetrahedral, octahedral, and icosahedral groups. One can easily check that our gate g cannot generate a representation of one of these groups, and hence densely generates some infinite group G. By the same reasoning as in Corollary 3.3.2, we conclude that G is a Lie subgroup of $SO(3)$.

The Lie subgroups of $SO(3)$ are $SO(3)$, $U(1)$ (all rotations about one axis) and $U(1) \times \mathbb{Z}_2$ (all rotations about one axis, plus a rotation by π perpendicular to the axis). Again one can easily eliminate the possibility that G is $U(1)$ or $U(1) \times \mathbb{Z}_2$, and hence G must be all of $SO(3)$.

We have proven universality on three modes for real nontrivial g with determinant $+1$. Universality on $m \geq 3$ modes follows by a real analogue of Reck *et al.* [208], namely that any rotation matrix in $SO(m)$ can be expressed as the product of $O(m^2)$ real 2×2 optical gates of determinant 1. □

3.4 Open Questions

At the moment our dichotomy theorem only holds for beamsplitters, which act on two modes at a time and have determinant -1. As we said before, we leave open whether the dichotomy can be extended to two-mode gates with determinant other than -1. Although the phases of gates are irrelevant in the qubit model, the phases unfortunately *are* relevant in linear optics—and that is the source of the difficulty. Note that the previous universality result of Reck *et al.* [208] simply assumed that arbitrary phaseshifters were available for free, so this issue did not arise.

Another open problem is whether our dichotomy can be extended to k-mode optical gates for all constants k. Such a result would complete the linear-optical analogue of the dichotomy conjecture for standard quantum circuits. The case $k = 3$ seems doable because the representations of all exceptional finite subgroups of $SU(4)$ are known [142]. But already the case $k = 4$ seems more difficult, because the representations of all finite subgroups of $SU(5)$ have not yet been classified. Thus, a proof for arbitrary k would probably require more advanced techniques in representation theory.

Chapter 4

Classifying Commuting Hamiltonians

In this chapter, we start making progress towards classifying the space "below BQP" defined by restricted gate sets on qubits in terms of their exact sampling hardness. To do so, we will start with the simplest sorts of gate sets imaginable - namely two-qubit commuting gate sets. As shown by Bremner, Jozsa, and Shepherd [72], some commuting gate sets - namely CCZ, CZ, Z conjugated by Hadamards or CZ, Z, T conjugated by Hadamards - are capable of sampling tasks which are hard to simulate to constant multiplicative error. In this chapter, we extend this to try to classify the power of all w-qubit commuting gate sets - which may be diagonal in arbitrary bases. For technical reasons, we will solve this in a slightly easier settings of commuting *Hamiltonians* rather than gates. Hamiltonians are continuous analogues of gates, so by considering Hamiltonians we can use mathematical methods for continuous or Lie groups in order to solve the classification. We give a complete classification of the power of commuting Hamiltonians - in particular any commuting Hamiltonian H which gives rise to entanglement can be used to perform hard sampling problems. This can be seen as progress towards Conjecture 2.2.1. We leave open the problem of classifying all commuting two-qubit gate sets.

This chapter is based on joint work with Laura Mančinska and Lucy Zhang [66].

4.1 Introduction

4.1.1 Problem statement and results

The evolution of a quantum system is determined by its Hamiltonian, which corresponds to a Hermitian matrix H. If we apply a Hamiltonian for time t, then this applies the unitary gate e^{iHt} to the system. The Hamiltonian of a system is governed by its underlying physics, so oftentimes in quantum computing experiments (e.g. in superconducting qubits) it is easy to apply certain Hamiltonians but not others. From this perspective it is natural to study the computational power of a fixed Hamiltonian H that can be applied to different ordered subsets of qubits for arbitrarily chosen amounts of time. Here we consider the model where we have a fixed two-qubit[1] Hamiltonian H which we can apply to any ordered pair of qubits, where we initialize our system in a computational basis state and perform a computational basis measurement at the end. Now it is natural to ask: What is the computational power of this model for a fixed H? It is known that almost any choice of H in this model yields universal quantum computation [102, 240, 90, 52], but the classification of such universal

[1] One-qubit Hamiltonians cannot create entanglement, so are efficiently classically simulable in this model.

Hamiltonians remains an open problem. Curiously, there exist subsets of Hamiltonians that do not seem to offer the full power of BQP but nevertheless are hard to simulate classically under plausible complexity assumptions [216, 72, 217].

In this chapter, we focus on a particular family of Hamiltonians H which, even though incapable of universal quantum computation, can perform computations that are hard for classical computers and might offer easier experimental implementation. Specifically, we study Hamiltonians H that can only give rise to mutually commuting gates, so the order in which the gates are applied is irrelevant:

Definition 4.1.1. We say that a two-qubit Hamiltonian H is *commuting* if $[H\otimes I, I\otimes H] = 0$ and $[H\otimes I, I\otimes (THT)] = 0$ and $[H, THT] = 0$, where T is the gate which exchanges two qubits, and $[A, B]$ denotes the quantity $AB - BA$. In other words, H commutes with itself when applied to any pair of qubits.

We are interested in classifying which commuting two-qubit Hamiltonians H allow us to perform computational tasks that are hard for classical computers. In particular, we want to understand when H gives rise to probability distributions which are hard to weakly simulate classically to constant multiplicative error. This can be seen as a simple subproblem of Conjecture 2.2.1.

Clearly, if a commuting H is not capable of creating entanglement from any computational basis state then the system will remain in a product state, so this model will be efficiently classically simulable. Surprisingly, we show that in all the remaining cases H can perform sampling tasks which cannot be simulated classically unless PH collapses.

Theorem 4.1.2 (Main Result). *If a commuting two-qubit Hamiltonian H is capable of creating entanglement from a computational basis state, then it gives rise to probability distributions that are hard to weakly simulate classically to constant multiplicative error unless* PH *collapses.*

Additionally, given such an H, our result provides an algorithm which describes the experimental setup required to sample from these hard distributions.

4.1.2 Proof ideas

Our proof proceeds in several steps. First, we use that fact that any commuting two-qubit Hamiltonian H is locally diagonalizable:

Lemma 4.1.3 ([97] Lemma 33). *For any commuting two-qubit Hamiltonian there exists a one-qubit unitary U and a diagonal matrix D such that $H = (U\otimes U)D(U^\dagger \otimes U^\dagger)$.*

The proof of this follows from expanding H in the Pauli basis, and deducing relationships between the Pauli coefficients.

Next, we use postselection gadgets to construct a family of one-qubit operations $L(t) : \mathbb{C}^2 \to \mathbb{C}^2$ for $t \in \mathbb{R}$ that that can be applied to the input state using postselection. We then show that these gadgets are universal on a qubit whenever H generates entanglement, so long as H is not some exceptional subcase. The exceptional subcase is $H = X(\theta)\otimes X(\theta)$ where $X(\theta) = \begin{pmatrix} 0 & e^{i\theta/2} \\ e^{-i\theta/2} & 0 \end{pmatrix}$.

Lemma 4.1.4. *If H is capable of creating entanglement from a computational basis state and H is not $X(\theta)\otimes X(\theta)$ for some θ, then it is possible to construct any one-qubit gate by taking products of the $L(t)$ gadgets.*

The main difficulty in proving this fact is that the maps $L(t)$ are in general *non-unitary*. Furthermore since they are generated with postselection, it is unclear how to invert them, so *a priori* they might not even form a group. Fortunately, we find new (and somewhat complicated) postselection gadgets to construct the L^{-1} operations, thus allowing us to apply group-theoretic and Lie-theoretic techniques to address this problem.

The rest of the proof follows from standard techniques in complexity, as discussed in Section 2.2.3. Since one-qubit gates plus any entangling Hamiltonian form a universal gate set [105, 71], our model can perform universal quantum computation under postselection.

Lemma 4.1.5. *If H is capable of creating entanglement from a computational basis state and H is not $X(\theta) \otimes X(\theta)$ for some θ, then postselected circuits involving H are universal for* BQP.

The proof of this statement uses a non-unitary version of the Solovay-Kitaev theorem proven by Aharonov et al. [29] to show our choice of gate set is irrelevant. Next, since we have a Solovay-Kitaev Theorem for these postselection gadgets, by Theorem 2.2.3 we see that postselecting our circuits further enables us to solve PP-hard problems. It then follows by Lemma 2.2.2 that a randomized classical algorithm cannot sample from the probability distributions produced by our circuits to constant multiplicative error unless the polynomial hierarchy collapses.

This completes the classification for all cases except the case $H = X(\theta) \otimes X(\theta)$. Hardness of sampling from these Hamiltonians was previously shown by Fefferman, Foss-Feig, and Gorshkov [112] using a construction which embeds permanents directly in the output distributions of such Hamiltonians. Hardness then follows from the arguments of Aaronson and Arkhipov [11]. We provide a summary of their hardness result for completeness.

4.1.3 Relation to prior work

As previously mentioned, this work is inspired by Bremner, Jozsa, and Shepherd [216, 72, 217], who showed that certain computations with commuting gates are hard to simulate classically unless the polynomial hierarchy collapses. In particular, they show hardness of simulating the gate set comprised of HZH, $H^{\otimes 2}CZH^{\otimes 2}$, and HPH, where P is the $\pi/8$-phase gate, as well as the gate set $H^{\otimes 3}CCZH^{\otimes 3}$, $H^{\otimes 2}CZH^{\otimes 2}$, HZH. Similarly, Shepherd [215, 217] considers the power of applying quantum Hamiltonians which are diagonal in the X basis, where the Hamiltonians can be applied only for discrete amounts of time θ; he describes the values of θ for which the resulting circuits are efficiently classically simulable or hard to weakly simulate (that is, to sample from the output probability distribution with a classical computer). Our work differs from these in several ways. First, We consider Hamiltonians rather than gates, and show hardness of *generic* or *average-case* commuting Hamiltonians, rather than showing hardness for worst-case commuting operations. Furthermore, we fully classify the computational complexity of all commuting Hamiltonians, and prove a dichotomy between hardness and classical simulability.

The hardness results we obtain in this chapter (as well as those in [72, 215, 217]) are based on the difficulty of sampling the output probability distribution on all n output qubits - also known as weak(n) simulation as discussed in section 2.2.2. A number of other works have considered the power of computations with commuting Hamiltonians, where one only considers the output distribution on a small number of output qubits - i.e. weak(1) or strong(1) simulation. For example, Bremner, Jozsa and Shepherd [72] showed that computing the marginal probability distributions on $O(\log(n))$ qubits of their model (i.e. strong($\log n$)

simulation) is in P. Ni and Van den Nest [198] showed that this holds for arbitrary 2-local commuting Hamiltonians, but also showed there exist 3-local commuting Hamiltonians for which this task is hard. Hence the problem of strongly simulating the output distributions (that is, being able to compute the probability of any event) of arbitrary k-local Hamiltonians is hard for $k \geq 3$. Along a similar line of thought, Takahashi et al. [228] showed that there is a system of 5-local commuting Hamiltonians for which weakly simulating the output on $O(\log(n))$ bits is hard.

As discussed in section 2.2.4 of this thesis, our work follows in the footsteps of a number of other authors who have considered "weak" models of quantum computation which can sample from difficult probability distributions.

Finally, other works have addressed the classification of universal two-qubit gates and Hamiltonians. Childs, Leung, Mančinska, and Ozols [90] classified the set of two-qubit Hamiltonians which give rise to $SU(4)$ when acting on two qubits, and are hence universal. Lloyd [180] and others [102, 240, 90, 52] have shown that a Haar-random two-qubit gate is universal with probability 1. Our work differs from these in that our Hamiltonians only become universal under postselection. Additionally, Cubitt and Montanaro [97] previously classified the complexity of two-qubit Hamiltonians in the Local Hamiltonian Problem setting. Specifically, given a two qubit Hamiltonian H, they classify the computational complexity of determining the ground state energy of Hamiltonians of the form $\sum_{ij} c_{ij} H_{ij}$ for real coefficients c_{ij}. This is incomparable with our classification, since we are studying the power of the Hamiltonian dynamics (in which the system is not in the ground state), rather than the complexity of their ground states.

4.2 Preliminaries and statement of Main Theorem

A two-qubit Hamiltonian H is a 4×4 Hermitian matrix. Let T denote the SWAP gate which exchanges two qubits, i.e.

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

so T maps the state $a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle$ to the state $a|00\rangle + c|01\rangle + b|10\rangle + d|11\rangle$. Given H, we assume that one can apply either H or THT to any pair of qubits. In other words, we can apply the Hamiltonian oriented from qubit i to qubit j, or from qubit j to qubit i. We will use H_{ij} to denote the Hamiltonian applied from qubit i to qubit j. Additionally, we will assume we can apply $-H$ as well, i.e., we can perform the inverse Hamiltonian.[2]

Suppose we are given some input string $x \in \{0,1\}^n$, and we want to define a distribution on $n' = \text{poly}(n)$ bits which we can efficiently sample from using H. Suppose we initialize a system of n' qubits in a computational basis state $|y\rangle$ for $y \in \{0,1\}^{n'}$, apply each Hamiltonian H_{ij} for time $t_{ij} \in \mathbb{R}$, and then measure all the qubits in the computational basis. (Here the times t_{ij} and the string y may depend on x.) This will induce some probability distribution

[2] If we had only assumed access to H and positive time evolution, we could always approximate the action of $-H$; this follows from compactness of the unitary group and was shown e.g. in Appendix A of [90]. However, here we are assuming we have exact access to $-H$; this will be useful when arguing about post-selected versions of these circuits.

$\mathcal{D}_x$ over bit strings of length n' on the output bits. Intuitively, these are the sorts of distributions one can efficiently sample from using H, using circuits which start and end in the computational basis.

However, this definition does not quite suffice to capture a realistic model of computation, because we have not specified how the initial state y and the times t_{ij} are chosen. To fix this, we will require that one could use a classical computer to efficiently calculate the experimental setup for each n. In other words, we will require that there exists a polynomial-time algorithm which, given $x \in \{0,1\}^*$, computes the values of y and t_{ij} used in the experiment. Furthermore, we will require that the times t_{ij} can be represented with polynomially many bits, and that they are all bounded in magnitude by a polynomial in n. This ensures that as the size of the system grows, the amount of time one needs to run the Hamiltonian does not grow too quickly. In complexity theory this is called a *uniformity* condition. This requirement ensures that any advantage over classical computation arising from this model comes from the power of the quantum computation performed, not the computation of the experimental setup.

This is stated more formally as follows:

Definition 4.2.1. Let samp-IQP(H) denote those families of probability distributions $\{\mathcal{D}_x\}$ for which there exists a classical poly-time algorithm $\mathcal{A}$ which, given an input $x \in \{0,1\}^n$, outputs the specifications for a quantum circuit using H whose output distribution family is $\{\mathcal{D}_x\}$. In particular, $\mathcal{A}$ specifies a number of qubits $n' = \text{poly}(n)$, a string $y \in \{0,1\}^{n'}$ and and a series of times $t_{ij} \in \mathbb{R}$, such that running a quantum circuit starting in the state $|y\rangle$, applying the operator $e^{it_{ij}H_{ij}}$ for each (i,j), and then measuring in the computational basis will yield a sample from $\mathcal{D}_x$. Each t_{ij} must be specifiable with $\text{poly}(n)$ bits and be bounded in magnitude by a polynomial in n.

In short, the class samp-IQP(H) captures the set of probability distributions one can efficiently sample from using H. In our work, we will show that a classical randomized algorithm cannot weakly simulate these distributions to constant multiplicative error (for a definition of this see section 2.2.2).

We can now more precisely state our Main theorem: that our commuting circuits cannot be weakly simulated unless the polynomial hierarchy PH collapses:

Theorem 4.2.2 (Main Theorem). *If H is capable of generating entanglement from the computational basis, then* BPP *machines cannot weakly simulate* samp-IQP(H) *with multiplicative error $c < 1/2$ unless* PH *collapses to the third level.*

In other words, there is a dichotomy: either computations which H are efficiently classically simulable, or else they cannot be efficiently simulated unless the polynomial hierarchy collapses. As the non-collapse of the polynomial hierarchy is a widely accepted conjecture in computational complexity, this is strong evidence that samp-IQP(H) circuits are not efficiently classically simulable.

We now proceed to a proof of the Main Theorem.

4.3 Proof of Main Theorem

The basic idea is to use postselection gadgets to show that postselected samp-IQP(H) circuits are capable of performing universal quantum computation. Hence, adding further postselections allows one to decide any language in PostBQP. By Lemma 2.2.2, this proves hardness of weakly simulating such circuits unless PH collapses.

Proof of Theorem 4.2.2. Suppose we have a commuting two-qubit Hamiltonian H. The first step in our proof is to characterize the structure of such H. It is clear that if H is diagonal under a local change of basis, i.e. $H = (U \otimes U)D(U^\dagger \otimes U^\dagger)$ for some one-qubit $U \in SU(2)$ and diagonal matrix D, then H is commuting. However, it is possible *a priori* that there exist commuting Hamiltonians which are not of this form. If T is the gate that swaps two qubits, then the fact that H is commuting implies that $H \otimes I$, $(THT) \otimes I$, $I \otimes H$, and $I \otimes (THT)$ are all simultaneously diagonalizable. However, it might be that this simultaneous diagonalization can only happen under a non-local change of basis. Fortunately, it turns out this is not possible - any commuting Hamiltonian must be locally diagonalizable. This was first shown by Cubitt and Montanaro [97], as previously noted in Lemma 4.1.3. We defer the proof of Lemma 4.1.3 to Section 4.4, which uses expansion in the Pauli basis. One can also prove this fact using linear algebra, but the proof becomes complicated in the case of degenerate eigenvalues. We thank Jacob Taylor for pointing us to this simplified proof, and Ashley Montanaro for pointing us to the proof in reference [97].

By Lemma 4.1.3, we know that $H = (U \otimes U)\,\mathrm{diag}(a, b, c, d)(U^\dagger \otimes U^\dagger)$ for some one-qubit unitary $U = \begin{pmatrix} \alpha & -\beta^* \\ \beta & \alpha^* \end{pmatrix}$ and some real parameters a, b, c, d. The trace of H contributes an irrelevant global phase to the unitary operator it generates, so without loss of generality we can assume H is traceless, i.e., $a + b + c + d = 0$.

Note that if $a = d = -1$, $b = c = 1$, and $|\alpha| = |\beta|$, then we have that $H = X(\theta) \otimes X(\theta)$, where $e^{i\theta} = \alpha/\beta$. As mentioned previously, these Hamiltonians are hard to simulate by an independent hardness result of Fefferman *et al.* [112], so in the rest of our proof, we will assume we are not in the case $a = d = -1$, $b = c = 1$ and $|\alpha| = |\beta|$. For completeness we will provide a summary of their work at the end of this proof.

We now consider the conditions under which computations with H are efficiently classically simulable. First, if H is diagonal in the computational basis, then it is obviously classically simulable, because it cannot generate entanglement from the computational basis. This corresponds to the case that $\alpha = 0$ or $\beta = 0$. So we can assume for the result of the proof that $\alpha \neq 0$ and $\beta \neq 0$.

Another way that H can fail to generate entanglement from the computational basis is if $b + c = a + d$. Since we are assuming the Hamiltonian is traceless this is equivalent to the condition $b + c = 0$. Indeed if H satisfies $b + c = 0$, and H is traceless so $a + d = 0$, then it is easy to check that e^{iHt} is nonentangling for all $t \in \mathbb{R}$. So we can assume in the rest of the proof that $b + c \neq 0$.

We now show that for all remaining H, we have $\mathsf{PostBQP} \subseteq \mathsf{PostIQP(H)}$. To do so, we break into two cases. Either $b = c$, so $H = THT$ and the Hamiltonian is identical when applied from qubit 1 to 2 vs. from 2 to 1, or $b \neq c$ so $H \neq THT$. For clarity of presentation, we will prove our main theorem in the case $b = c$, as this proof uses simpler notation. An analogous proof holds for the case $b \neq c$, which we defer to Section 4.7.

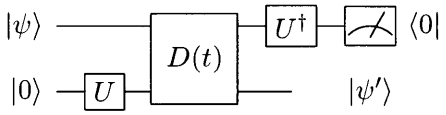
Now in the case $b = c$, consider the rescaled Hamiltonian $H' = H/b$. Since $b + c \neq 0$ and $b = c$ this Hamiltonian is well-defined, and we have $H' = (U \otimes U)\,\mathrm{diag}(a', 1, 1, d')(U^\dagger \otimes U^\dagger)$ for some real parameters a' and d' which obey $a' + d' = -2$. Now consider the two-qubit unitary $V(t)$ we obtain from running H' for time $t \in \mathbb{R}$

$$V(t) = e^{itH'} = (U^{\otimes 2})D(t)(U^{\dagger \otimes 2}),$$

where $D(t) \triangleq \mathrm{diag}(e^{ia't}, e^{it}, e^{it}, e^{id't})$. Here we have used the fact that if U is an arbitrary unitary, then $e^{UH'U^\dagger} = Ue^{H'}U^\dagger$.

A samp-IQP(H) circuit is specified by times t_{ij} for all unordered[3] pairs of qubits (i,j), as well as an initial basis state $|y\rangle$ for $y \in \{0,1\}^{\text{poly}(n)}$. The circuit consists of applying $V(t_{ij})$ to each pair of qubits (i,j) to $|y\rangle$, and then measuring in the computational basis. This can be easily seen to be equivalent to the following circuit: Start in the state $|y\rangle$, apply U to every qubit, then apply $D(t_{ij})$ to each pair of qubits; finally, apply $U^\dagger$ to every qubit and measure in the computational basis. (This is true because all factors of U and $U^\dagger$ in the circuit cancel except those at the beginning and end).

We will now show how to make post-selected gates of this form perform universal quantum computing. The basic idea is that we already have a two-qubit entangling Hamiltonian at our disposal. Therefore, if we could show how to perform arbitrary one-qubit gates using post-selection, this would form a universal gate set for quantum computing by the result of Dodd et al. [105] or Bremner et al. [71]. Following the method of Bremner, Jozsa, and Shepherd [72], we consider the following post-selection gadget, denoted $L(t)$, which performs an operation on a single qubit state $|\psi\rangle$:



The postselection is denoted in the circuit by $\langle 0|$. Note that this gadget preserves the property that every line begins with $U|0\rangle$, and ends with $U^\dagger$ and a measurement. Hence, if we could use these postselection gadgets to perform arbitrary single-qubit gates, then we could perform universal quantum computing under postselection as follows: Given a target quantum circuit to simulate, compile the circuit out of gates of the form $D(t)$ and single-qubit gates. Additionally, add a $UU^\dagger$ (which is the identity) at the beginning and end of every line, so that each line starts with a U and ends with a $U^\dagger$. Now this circuit consists of applying a column of U's, then a series of diagonal gates $D(t)$ and one-qubit gates, followed by a column of $U^\dagger$s. This almost has the form of a samp-IQP(H) circuit, with the exception of the one-qubit gates (note that these include both the gates $U^\dagger$ in the second column and the gates U in the second to last column). Now for each one-qubit gate g, replace it with its implementation using postselection gadgets $L(t)$. After this transformation, each line begins with a U, ends with a $U^\dagger$, and contains only diagonal gates $D(t)$ in the interior of the circuit. However, now we've additionally specified some postselection bits, so we have created a PostIQP(H) circuit which simulates universal quantum computing.

Let us examine what transformation $L(t)$ actually performs on the qubits involved. The gadget performs some linear transformation on the input state $|\psi\rangle$. In particular, it acts on $|\psi\rangle$ by

$$L(t) = \frac{1}{|\alpha||\beta|\sqrt{-2i\sin(2t)}} \begin{pmatrix} |\alpha|^2 e^{ia't} & \alpha\beta^* e^{it} \\ \alpha^*\beta e^{it} & |\beta|^2 e^{id't)} \end{pmatrix}.$$

This is a non-unitary transformation, so it does not preserve the norms of vectors. Since we only care about how $L(t)$ behaves on the projective Hilbert space of quantum states, we can choose the overall normalization so that $L(t) \in SL(2,\mathbb{C})$. Note that this operator is well-defined only if the denominator above is non-zero, so we will require that $t \in (0,\pi) \cup (\pi, 2\pi)$.

In addition to being able to perform the transformation $L(t)$ as t ranges over $t \in (0,\pi) \cup (\pi, 2\pi)$, we can also perform products of such transformations. In fact, we can perform any

[3]This is because we are considering the case $b = c$ i.e. $H = THT$.

operation in the set

$$S \triangleq \overline{\langle \{L(t) : t \in (0,\pi) \cup (\pi, 2\pi)\} \rangle}.$$

Here the angled brackets $\langle A \rangle$ denote the set of all matrices obtained by finite products of elements of A. The bar above $\overline{\langle A \rangle}$ means that we take the closure of this set in $SL(2,\mathbb{C})$; in other words, we include all matrices that one can obtain by taking limits of sequences of finite products of A, so long as the limit point belongs to $SL(2,\mathbb{C})$.

If the matrices $L(t)$ were in a compact space such as $SU(2)$, then it would immediately follow that S contains inverses of all its elements.[4] Therefore we would know that S is a group, and we could apply tools from group theory to categorize S. However, our matrices are in the non-compact space $SL(2,\mathbb{C})$. Therefore it is not clear whether S is closed under taking inverses, so S *might not be a group*! Furthermore, since L is obtained under post-selection, the assumption that we can perform the inverse of H does not imply we can perform L^{-1}.

To fix this problem, we find additional gadgets which allow us to construct L^{-1} by adding additional postselections to our circuit. In particular, we will show that for each $L(t)$, there exists a postselection gadget of finite size which performs $L(t)^{-1}$ exactly. An important restriction on this construction is that this inverse must be efficiently computable. Specifically, for each $L(t)$ the size of the postselection gadget required to invert $L(t)$ is of constant size. Additionally, the construction of the postselection gadget will in general contain several time parameters which one needs to set in order to obtain $L(t)^{-1}$. We also require that we can set these times so that we obtain L^{-1} to accuracy ε in $\text{polylog}(k_L 1/\varepsilon)$ time, where k_L is a constant which depends on $L(t)$ only. Furthermore, the amount of time needed to run the Hamiltonians in the inverse gadget are bounded above by a polynomial. For convenience we will refer to these properties as "the construction is efficiently computable."

At first glance it might sound like this definition of "efficiently computable" is too weak, because the inverses of arbitrary L matrices might require large postselection gadgets. However, later in our construction we will use the fact that for any fixed Hamiltonian H, we will only need to invert a finite set of L matrices. Hence for fixed H, the size of the postselection gadgets which appear in our circuit will be upper bounded by a constant depending on H only, but not on the size of the problem we are solving under postselection. Furthermore, for fixed H, we can compute the times in the inversion gadgets to invert the relevant L matrices to exponential accuracy in polynomial time. This ability to invert the L matrices to exponential precision will later be crucial for our hardness of sampling result.

Furthermore, note that in the case that $H \neq THT$, the construction of these gadgets can be made substantially simpler. In particular, the gadgets to construct $L^{-1}(t)$ are of size 4 for any t, and the times used in running the Hamiltonians are trivially efficiently computable to polynomial digits of accuracy. From a practical experimental perspective these circuits would be easier to construct, and since $H \neq THT$ is the generic case for commuting Hamiltonians, would be applicable for almost all commuting Hamiltonians. We include this construction in Section 4.7 for the interested reader.

Claim 4. *For any given $L(t)$, where $t \in (0,\pi) \cup (\pi, 2\pi)$, it is possible to construct $L(t)^{-1}$ by introducing a constant number of postselections and a constant number of ancillae into*

[4] To see this, take an element $s \in S$. If s has finite order, than its inverse is clearly in S. If s has infinite order, consider the sequence $1, s, s^2, \ldots$. Since the matrices are in a compact space T, the sequence of powers must have a convergent subsequence, i.e. there must be positive $n_1, n_2, n_3 \ldots$ such that $n_1 < n_2 < \ldots$ and $s^{n_1}, s^{n_2}, \ldots$ approach some element $t \in T$. Therefore the sequence $s^{n_2-n_1}, s^{n_3-n_2}, \ldots$ must approach the identity, and the sequence $s^{n_2-n_1-1}, s^{n_3-n_2-1}, \ldots$ must approach s^{-1}.

the circuit. Furthermore, this construction is efficiently computable in the manner described above.

The proof of Claim 4 can be found in Section 4.5, and is somewhat involved.

We now redefine S so that its base set contains these inverses:

$$S \triangleq \overline{\langle \{L(t) : t \in (0,\pi) \cup (\pi, 2\pi)\} \cup \{L(t)^{-1} : t \in (0,\pi) \cup (\pi, 2\pi)\} \rangle}.$$

Using this definition, we can now show using standard techniques that S is a Lie group—this is essentially a consequence of Cartan's closed subgroup theorem [83] and the fact that inversion is a continuous operation in the matrix entries on $SL(2,\mathbb{C})$. Once we know that S has the structure of a Lie group, we can apply the theory of Lie algebras to identify what set of matrices are in S. In particular, we can show that S generates all of $SL(2,\mathbb{C})$.

Claim 5. $S = SL(2,\mathbb{C})$.

The proof of this claim is a tedious but straightforward calculation using Lie algebras and properties of the exponential map on $SL(2,\mathbb{C})$. The proof uses the fact that we are not in one of the cases excluded by our theorem (i.e. H does generate entanglement and is not $X(\theta) \otimes X(\theta)$ for some θ) - in these cases one does *not* find that $S = SL(2,\mathbb{C})$ as one would expect. In certain special cases, the gadgets $L(t)$ alone do not generate $SL(2,\mathbb{C})$, specifically when $a' = \pm 1$ or $a' = -3$. In these cases, we show that one can add additional postselection gadgets, which are closed under taking inverses, which boost the power of the $L(t)$ transformations to cover all of $SL(2,\mathbb{C})$. This simply reflects that for very particular Hamiltonians, our L matrices need additional help to span all 1-qubit operations. We include the proof in Section 4.6.

Now that we have shown density in $SL(2,\mathbb{C})$, as well as the fact that we can produce inverses of the gates in our generating set, our proof of yielding PP under postselection follows almost immediately. In particular, we will invoke the following theorem by Aharanov, Arad, Eban and Landau [29]:

Theorem 4.3.1 ([29] Theorem 7.6, adapted to our case)**.** *There exists a constant $\varepsilon_0 > 0$ such that, for any $G = \{g_1 \dots g_k\} \subset SL(2,\mathbb{C})$ which is an ε_0-net over B, where B is the set of operations in $SL(2,\mathbb{C})$ which are 2.1-far from the identity (which in particular contains $SU(2)$), then for any unitary $U \in SU(2,\mathbb{C})$, there is an algorithm to find an ε-approximation to U using* polylog$(1/\varepsilon)$ *elements of G and their inverses which runs in* polylog$(1/\varepsilon)$ *time.*

In the above theorem, when we say an operation is "ε-far" from another, we are referring to the operator norm.

From this, we can immediately prove the main theorem. Suppose we wish to compute a language $L_0 \in$ PP, and we have a commuting Hamiltonian H of the form promised in Theorem 4.2.2. By Aaronson's result that PP $\subseteq$ PostBQP [6], there is an efficiently computable postselected quantum circuit C composed of Hadamard and Toffoli gates which computes L. Additionally, by Claim 5 there exists a finite set G of products of L's and L^{-1}'s which form an ε_0-net over B (which can be computed in finite time). Hence by Theorem 4.3.1 there is a poly-time algorithm which expresses single-qubit gates as products of elements of G to exponential accuracy. Likewise, since H is entangling, we can generate some entangling two-qubit gate g, as well as its inverse g^{-1} (by applying $-H$). Since g and single-qubit gates are universal [71], by the usual Solovay–Kitaev theorem [99], we can express the circuit C in terms of g, g^{-1}, and single-qubit gates to exponential accuracy with polynomial overhead. Combining these, we can express the circuit C as a polynomial sized

product of g's, g^{-1}'s, L's, and L^{-1}'s , which we can express as a PostIQP(H) circuit using the gadgets described previously. Hence this PostIQP(H) circuit decides the language L_0.

Note that in this construction, it is crucial that we only ever need to invert a finite number of $L(t)$ matrices. This ensures that the size of the postselection gadgets involved to construct the L^{-1} operations are upper bounded by a constant depending on the choice of H only. Additionally, it is important that we can construct the L^{-1} matrices exponential accuracy. This is crucial because in order to perform PostBQP under postselection, one needs to be able to simulate Aaronson's algorithm to exponential accuracy [5]. Fortunately our construction allows us to simulate the algorithm to high accuracy, and hence these Hamiltonians can be used to sample from probability distributions which are not possible to simulate with a classical computer unless the polynomial hierarchy collapses.

This completes the proof in all cases except the exceptional case $H = X(\theta) \otimes X(\theta)$. This has a separate hardness of sampling result which was shown by Fefferman, Foss-Feig, and Gorshkov [112]. In particular, they showed the following:

Theorem 4.3.2 (Fefferman et al. [112]). *If $H = X(\theta) \otimes X(\theta)$ for some θ, then a* BPP *machine cannot weakly simulate* samp-IQP(H) *with any constant multiplicative error unless* PH *collapses to the third level.*

Their proof makes use of that fact that using such Hamiltonians, for any matrix $A \in \{0, \pm 1\}^n$, one can perform a unitary U on a system of $O(n)$ qubits such that $\langle 1^n|U|0^n\rangle = k\,(\mathrm{Perm}(A) + \varepsilon)$, where k is independent of A and exponentially small in n, $\mathrm{Perm}(A)$ denotes the permanent of A, and ε is a term with norm $o(2^{-n})$. Note that $\mathrm{Perm}(A)^2$ is #P-hard to compute with any constant multiplicative error [11]: Therefore Theorem 4.3.2 immediately follows by the techniques of Aaronson and Arkhipov [11] - because if there were an efficient classical simulation of such circuits, then using approximate counting [225], one could approximate $\mathrm{Perm}(A)^2$ to multiplicative error $\left(1 + \frac{1}{\mathrm{poly}(n)}\right)$ in $\mathsf{BPP}^{\mathsf{NP}}$. But $\mathsf{BPP}^{\mathsf{NP}} \subseteq \Delta_3$, so again by Toda's theorem [232] this implies the collapse of PH to the third level.

This completes the last remaining case, and hence completes the proof. We provide complete proofs of all Lemmas not shown here in the following sections.

□

4.4 Commuting Hamiltonians are locally diagonalizable

To establish Lemma 4.1.3, we prove the following stronger statement.

Claim 6. *If H is a two-qubit Hamiltonian and $[H \otimes I, I \otimes H] = 0$, then $(U \otimes U)H(U \otimes U)^\dagger$ is diagonal for some one-qubit unitary U.*

This is actually slightly stronger than Lemma 33 of [97], which shows that if $[H \otimes I, I \otimes H] = [H \otimes I, I \otimes THT] = [THT \otimes I, I \otimes H] = 0$, then H is locally diagonalizable. Here we merely require that $[H \otimes I, I \otimes H] = 0$.

Proof. As a first step we expand H in Pauli basis and let α_{AB} be the coefficient at $A \otimes B$ term for any $A, B \in \{I, X, Y, Z\}$. Also, for all $A \in \{I, X, Y, Z\}$, let

$$\vec{c}_A := (\alpha_{XA}, \alpha_{YA}, \alpha_{ZA})^T \quad \text{and} \quad \vec{r}_A := (\alpha_{AX}, \alpha_{AY}, \alpha_{AZ})^T. \tag{4.1}$$

[5] This is because the algorithm postselects on an exponentially unlikely event, so to maintain polynomial accuracy after postselection we require exponential accuracy prior to postselection.

Given a vector $\vec{v} = (v_x, v_y, v_z)^T \in \mathbb{R}^3$, we adopt a commonly used notation and write $\vec{v} \cdot \vec{\sigma}$ to denote the linear combination $v_x X + v_y Y + v_z Z$.

Since $(H \otimes I)(I \otimes H) = (I \otimes H)(H \otimes I)$, we know that both products must have the same expansion in Pauli basis. Let us fix $A, B \in \{I, X, Y, Z\}$ and consider the terms of the form $A \otimes \_ \otimes B$ in the Pauli expansion of each of the products.

First, for $(H \otimes I)(I \otimes H)$ we notice that, when restricted to terms of the form $A \otimes \_ \otimes B$, its Pauli expansion is given by

$$\big(A \otimes (\alpha_{AI} I + \vec{r}_A \cdot \vec{\sigma}) \otimes I\big)\big(I \otimes (\alpha_{IB} I + \vec{c}_B \cdot \vec{\sigma}) \otimes B\big) = \tag{4.2}$$

$$A \otimes \big(\alpha_{AI}\alpha_{IB} I + (\alpha_{AI}\vec{c}_B + \alpha_{IB}\vec{r}_A) \cdot \vec{\sigma} + (\vec{r}_A \cdot \vec{\sigma})(\vec{c}_B \cdot \vec{\sigma})\big) \otimes B = \tag{4.3}$$

$$A \otimes \big((\alpha_{AI}\alpha_{IB} + \vec{r}_A \cdot \vec{c}_B) I + (\alpha_{AI}\vec{c}_B + \alpha_{IB}\vec{r}_A + i(\vec{r}_A \times \vec{c}_B)) \cdot \vec{\sigma}\big) \otimes B, \tag{4.4}$$

where we have applied the identity $(\vec{v} \cdot \vec{\sigma})(\vec{w} \cdot \vec{\sigma}) = (\vec{v} \cdot \vec{w}) I + i(\vec{v} \times \vec{w})\vec{\sigma}$ in the last step.

Next, we consider the product $(I \otimes H)(H \otimes I)$ and similarly obtain that, when restricted to terms of the form $A \otimes \_ \otimes B$, its the Pauli expansion is given by

$$\big(I \otimes (\alpha_{IB} I + \vec{c}_B \cdot \vec{\sigma}) \otimes B\big)\big(A \otimes (\alpha_{AI} I + \vec{r}_A \cdot \vec{\sigma}) \otimes I\big) = \tag{4.5}$$

$$A \otimes \big((\alpha_{AI}\alpha_{IB} + \vec{c}_B \cdot \vec{r}_A) I + (\alpha_{AI}\vec{c}_B + \alpha_{IB}\vec{r}_A + i(\vec{c}_B \times \vec{r}_A)) \cdot \vec{\sigma}\big) \otimes B. \tag{4.6}$$

Since the coefficients in the Pauli expansions of $(H \otimes I)(I \otimes H)$ have to coincide with those in the expansion of $(I \otimes H)(H \otimes I)$, we know that the difference between expressions (4.4) and (4.6) equals zero. Considering the middle tensor and canceling some therms gives

$$(\vec{r}_A \times \vec{c}_B) \cdot \vec{\sigma} = (\vec{c}_B \times \vec{r}_A) \cdot \vec{\sigma}. \tag{4.7}$$

Since $\vec{v} \times \vec{w} = -\vec{w} \times \vec{v}$, we obtain that $\vec{r}_A \times \vec{c}_B = 0$. This further implies that $\vec{r}_A$ and $\vec{c}_B$ are collinear, that is, $\dim(\mathrm{span}\{\vec{r}_A, \vec{c}_B\}) \leq 1$. Since we can choose arbitrary $A, B \in \{I, X, Y, Z\}$, it must be that all the vectors $\vec{r}_A$ and $\vec{c}_B$ must lie in the same one-dimensional subspace, i.e.,

$$\dim\big(\mathrm{span}\{[\}]\vec{r}_A, \vec{c}_B : A, B \in \{I, X, Y, Z\}\big) \leq 1. \tag{4.8}$$

Let us now consider a 3×3 matrix M whose rows and columns are indexed by Pauli matrices X, Y and Z and its entries are defined via $M_{AB} = \alpha_{AB}$. Then the vectors $\vec{c}_A$ are the columns of M and $\vec{r}_B$ are its rows. From Equation (4.8), we see that M has rank at most one. Moreover, the row and column spaces of M must coincide as

$$\mathrm{span}\left(\{\vec{r}_X, \vec{r}_Y, \vec{r}_Z\}\right) = \mathrm{span}\left(\{\vec{c}_X, \vec{c}_Y, \vec{c}_Z\}\right). \tag{4.9}$$

These two observations imply that $M = \vec{v}\vec{v}^T$ for some $\vec{v} \in \mathbb{R}^3$. So we can express our Hamiltonian H as

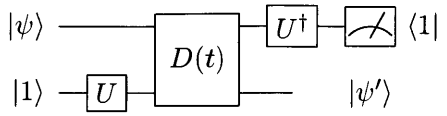
$$H = \alpha_{II} I \otimes I + (a\vec{v} \cdot \vec{\sigma}) \otimes I + I \otimes (b\vec{v} \cdot \vec{\sigma}) + (\vec{v} \cdot \vec{\sigma}) \otimes (\vec{v} \cdot \vec{\sigma}), \tag{4.10}$$

where $a, b \in \mathbb{R}$ are such that $\vec{r_I} = a\vec{v}$ and $\vec{c_I} = b\vec{v}$. If we pick a unitary U that diagonalizes $\vec{v} \cdot \vec{\sigma}$, then from Equation (4.10) we see that $U \otimes U$ diagonalizes our Hamiltonian H. This concludes the proof. □

4.5 Inverting L matrices using postselection gadgets

We now prove Claim 4.

Proof. We will need two additional gadgets for our construction. First, consider a modification of the gadget for $L(t)$, where we start the qubit in the $|1\rangle$ state and postselect on the $|1\rangle$ state:



By a direct calculation, one can show the linear transformation performed on $|\psi\rangle$ is given by

$$M(t) = \frac{1}{|\alpha||\beta|\sqrt{e^{-2it} - e^{2it}}} \begin{pmatrix} |\beta|^2 e^{ia't} & -\alpha\beta^* e^{it} \\ -\alpha^*\beta e^{i} & |\alpha|^2 e^{id't} \end{pmatrix}$$

This is tantalizingly close to the inverse of L, which is

$$L(t)^{-1} = \frac{1}{|\alpha||\beta|\sqrt{e^{-2it} - e^{2it}}} \begin{pmatrix} |\beta|^2 e^{id't} & -\alpha\beta^* e^{it} \\ -\alpha^*\beta e^{it} & |\alpha|^2 e^{ia't} \end{pmatrix}$$

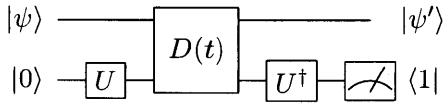
The only thing that is off is that the phase of the upper left and bottom right entries are incorrect. We now break into three cases to describe how to correct the phases in each. (Recall that $d' = -2 - a'$ as our without loss of generality our Hamiltonian is traceless).

Case 1: $a' = d' = -1$ In this case we already have $M(t) = L^{-1}(t)$, so we have found the inverse.

Case 2: $a' = 1, d' = -3$ OR $a' = -3, d' = 1$

We will prove the case $a' = 1$; an analogous proof holds for $a' = -3$.

To correct the phases in $M(t)$, we need to introduce an additional gadget:



In other words, instead of using the gate in a teleportation-like protocol, we instead use it to apply phases to $|\psi'\rangle$. This gate performs the following transformation on the input state:

$$N(t) = \frac{1}{\sqrt{(e^{it} - e^{ia't})(e^{id't} - e^{it})}} \begin{pmatrix} e^{it} - e^{ia't} & 0 \\ 0 & e^{id't} - e^{it} \end{pmatrix}$$

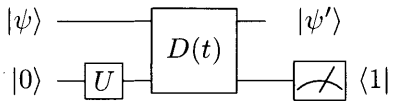
In the case that $a' = 1$, this gadget becomes singular, and hence it performs the operation

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

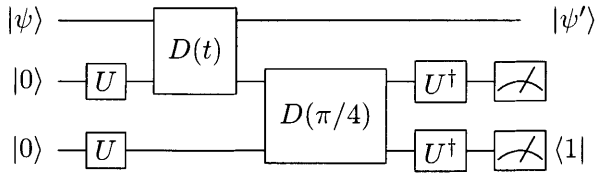
In other words, this gadget postselects the qubit involved on the state $|1\rangle$. This holds in particular for $t = \pi/4$. (In fact it holds for any t such that $e^{-3it} \neq e^{it}$, in which case it becomes undefined).

By composing $N(\pi/4)$ with other gadgets, this now empowers us to create gadgets in which we postselect on $|1\rangle$ on lines which do not end in $U^\dagger$. For instance, we can create the

following gadget:



Which one can easily check is equivalent to the following circuit, which maintains the property that every line begins and ends with U and $U^\dagger$.



This is simply composing the gadget with $N(\pi/4)$. (Here the output of the middle qubit is an independent sample from measuring the state $U^\dagger|1\rangle$ in the computational basis).

This gadget performs the following operation on $|\psi\rangle$:

$$P(t) \propto \begin{pmatrix} e^{it} & 0 \\ 0 & e^{-3it} \end{pmatrix} \propto \begin{pmatrix} e^{2it} & 0 \\ 0 & e^{-2it} \end{pmatrix}$$

In other words, the matrix $P(t)$ is a phase gate by phase $\theta = 2t$.

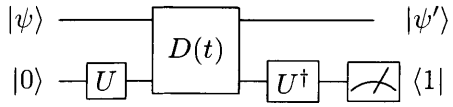
The construction of arbitrary phase gates suffices to correct the diagonal phases of $M(t)$, because for any matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we have that

$$\begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix} = \begin{pmatrix} ae^{i\theta} & b \\ c & de^{-i\theta} \end{pmatrix}$$

Hence by choosing $\theta = (d' - a')t$, and multiplying $M(t)$ by this matrix on both sides, we obtain $L^{-1}(t)$ as desired. Clearly this construction is efficient, i.e. the postselection gadget is of constant size, and one can efficiently compute the times to run the Hamiltonians in the gadget to high precision. This completes the proof.

Case 3: $a' \neq \pm 1, -3$

To correct the phases in $M(t)$, we need to consider the same gadget $N(t)$ which we used in Case 2:



In other words, instead of using the gate in a teleportation-like protocol, we instead use it to apply phases to $|\psi'\rangle$. This gate performs the following transformation on the input state:

$$N(t) = \frac{1}{\sqrt{(e^{it} - e^{ia't})(e^{id't} - e^{it})}} \begin{pmatrix} e^{it} - e^{ia't} & 0 \\ 0 & e^{id't} - e^{it} \end{pmatrix}$$

Since N is a diagonal matrix, the only physical quantity that matters is the ratio $r(t)$ of its

two entries, which is a complex number given by

$$r(t) = \frac{e^{it} - e^{ia't}}{e^{id't} - e^{it}}.$$

If $r(t)$ takes on a certain value, then it immediately follows that $N(t) = \pm \begin{pmatrix} \sqrt{r} & 0 \\ 0 & \sqrt{r}^{-1} \end{pmatrix}$, because of our normalization. Furthermore, if we compose $N(s)N(t)$, then the ratio of the resulting diagonal matrix is $r(s)r(t)$. Note the ± 1 term is an irrelevant global phase, so we omit it in the further calculations.

We will now show that for any complex phase $e^{i\theta}$, where $\theta \neq 0, \pi$, there exists a finite set of times $t_1, t_2, ...t_k$, $s_1, s_2, ...s_{k'}$ such that

$$N(t_1)N(t_2)...N(t_k)N(s_1)N(s_2)...N(s_{k'}) = \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix}$$

As previously mentioned in Case 2, the construction of such matrices suffices to correct the diagonal phases of $M(t)$, because for any matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we have that

$$\begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix} = \begin{pmatrix} ae^{i\theta} & b \\ c & de^{-i\theta} \end{pmatrix}$$

Hence by choosing $\theta = (d' - a')t$, and multiplying $M(t)$ by this matrix on both sides, we obtain $L^{-1}(t)$ as desired and this will complete the proof.

To prove this, we will prove two separate facts. First, we will show that given θ, there exists a sequence $t_1, t_2, ...t_k$ such that $N(t_1)N(t_2)...N(t_k) = \begin{pmatrix} ce^{i\theta/2} & 0 \\ 0 & \frac{1}{c}e^{-i\theta/2} \end{pmatrix}$ for some $c \in \mathbb{R}^+$. Next, we will show that for any $c \in \mathbb{R}$, there exists a sequence $s_1, s_2, ...s_{k'}$ of times such that $N(s_1)N(s_2)...N(s_{k'}) = \begin{pmatrix} 1/c & 0 \\ 0 & c \end{pmatrix}$. Together these imply the claim.

Moreover, we will show this construction is efficiently computable. More specifically, suppose you want to find invert L. The for each L the size of the postselection gadget required to invert L is of constant size. Additionally, the amount of computational time required to compute the values of t_i and s_i to ensure that we find L^{-1} to accuracy ε scales as polylog$(k_L * 1/\varepsilon)$, where k_L is a constant which depends on L. Furthermore, the times t_i and s_i are upper bounded by a constant which only depends on the value of a'. For convenience we will refer to these properties as "the construction is efficiently computable."

At first glance it might sound like this definition of "efficiently computable" is too weak, because the inverses of arbitrary L matrices might require large postselection gadgets, or might require a long time to compute the values of the t_i and s_i to sufficient accuracy. However, later in our construction we will use the fact that for any fixed Hamiltonian H, we will only need to invert a fixed number of L matrices. Hence for fixed H, the size of the postselection gadgets which appear in our circuit will be upper bounded by a constant depending on H only, but not on the size of the problem we are solving under postselection. Furthermore, for fixed H, we can compute the times t_i, s_i required to invert the relevant L matrices to exponential accuracy in polylog(1/*epsilon*) time (where a hidden constant k_L depending on L has been absorbed into the big-O notation).

Claim 7. *For any $\theta \in (0, 2\pi)$, there exists a sequence $t_1, t_2, ...t_k$ such that*

$$N(t_1)N(t_2)...N(t_k) = \begin{pmatrix} ce^{i\theta/2} & 0 \\ 0 & e^{-u\theta/2}/c \end{pmatrix}$$

for some $c \in \mathbb{R}^+$. Furthermore, this construction is computationally efficient.

Proof. To see this, consider the expression for the ratio

$$r(t) = \frac{e^{it} - e^{ia't}}{e^{id't} - e^{it}} = -\frac{1 - e^{i(a'-1)t}}{1 - e^{i(d'-1)t}}.$$

Let Phase(c) denote the phase of c modulo 2π. Then by direct calculation we have that

$$\begin{aligned}
\text{Phase}(r(t)) &= \pi + \text{Phase}\left(\frac{1 - e^{i(a'-1)t}}{1 - e^{i(-3-a')t}}\right) \\
&= \pi + \text{Phase}\left(1 - e^{i(a'-1)t}\right) - \text{Phase}\left(1 - e^{i(-3-a')t}\right) \\
&= \pi + \text{Phase}\left(1 - e^{i(a'-1)t}\right) + \text{Phase}\left(1 - e^{i(3+a')t}\right) \\
&= \left(\pi + \left(\frac{(a'-1)t}{2} \mod \pi\right) + \left(\frac{(3+a')t}{2} \mod \pi\right)\right) \mod 2\pi \\
&= \left(\pi + \left(t' \mod \pi\right) + \left(Rt' \mod \pi\right)\right) \mod 2\pi
\end{aligned}$$

Where $t' = (a'-1)t/2$ and $R = \dfrac{(3+a')}{(1-a')}$. Since we are in the case that $a' \neq \pm 1, -3$, we are promised that R is well-defined and $R \neq 0, 1$. Also note that we cannot have that $R = -1$ because this would imply $3 = -1$, a contradiction.

Suppose $R > 0$ (an analogous proof holds for $R < 0$). Then for $t' \in [0, \min(\pi, \pi/R)]$, we know that $\text{Phase}(r(t')) = \pi + (R+1)t'$, because in this range t' is sufficiently small such that both $t' \mod \pi = t'$ and $Rt' \mod \pi = Rt'$. Hence using t' in this interval, we can achieve any phase in $(\pi, \pi + s)$ where $s = (R+1)\min(\pi, \pi/R)$. For any $R \neq 0, -1$ this range is of constant size. Thus by multiplying together $1/s$ phases in the range $(\pi, \pi + s)$, one can achieve any phase in $(0, 2\pi)$, as desired.

Note that this construction is manifestly efficient; the t_i's are upper bounded by a constant $\min(\pi, \pi/R)$ which is a function of H only, and computing them to polynomially many digits requires polynomial time, as it just requires simple addition.

□

Claim 8. *For any $c \in \mathbb{R}^+ - \{1\}$, there exists a finite sequence $s_1, s_2, ...s_k$ such that*

$$N(s_1)N(s_2)...N(s_k) = \begin{pmatrix} 1/c & 0 \\ 0 & c \end{pmatrix}$$

Proof. Consider products of matrices of the form $N(s)N(-s)$ for $s \in \mathbb{R}+$. Let $f(s) = r(s)r(-s)$. One can check by direct calculation that

$$f(s) = \frac{1 - \cos((1-a')s)}{1 - \cos((3+a')s)}$$

In other words, the product of the ratios is real and positive, hence the resulting matrix $N(s)N(-s)$ is of the form $\begin{pmatrix} 1/\ell & 0 \\ 0 & \ell \end{pmatrix}$ for some $\ell \in \mathbb{R}^+$. Note since we are in the case $a' \neq \pm 1, -3$ this ratio is well-defined.

If we redefine $s' = s/(1-a')$, and set $R = (1-a')/(3+a')$, then this ratio becomes

$$\frac{1-\cos s'}{1-\cos Rs'}$$

We know $R \neq 0, 1$ because we have $a' \neq \pm 1, 3$, and furthermore $R \neq -1$ as well, since this would imply $1 = -3$, a contradiction.

For clarity of explanation assume $R > 0$; an analogous proof holds for the case $R < 0$.

Next we claim that the range of $f(s)$ as s varies over R includes the interval

$$(\min(R^{-2}, R^2), \max(R^{-2}, R^2)).$$

Since $R \neq 1$ this is an interval of constant size around 1. To see this, we will break into two cases.

First, assume $R > 1$. Consider the value of this function when $s' \in (0, \pi/R)$. The function $f(s')$ in continuous in this range. Additionally $\lim_{s' \to 0} f(s') = 1/R^2$ by L'Hôpital's rule, and $\lim_{s' \to \pi/R} = +\infty$. Hence the range of f covers $(R^{-2}, +\infty) = (\min(R^{-2}, R^2), +\infty)$ by the mean value theorem.

Next, assume $0 < R < 1$. Now consider the value of the function when $s' \in (0, \pi)$. Again the function is continuous in this range, and we have $\lim_{s' \to 0} f(s') = 1/R^2$ by L'Hôpital's rule, and $\lim_{s' \to \pi} = 0$. Hence the range of f covers $(0, R^{-2}) = (0, \max(R^{-2}, R^2))$ by the mean value theorem.

Hence in either case, by choosing an appropriate value of s', we can set $f(s)$ to be any real value in a finite-length interval containing 1. Hence for any target ratio $c^2 \in \mathbb{R}^+$, one can take a finite product of $O(\log(c))$ values of $f(s)$ such that $f(s_1)f(s_2)...f(s_k) = c^2$. This implies the claim.

Note that this construction is efficient. First, the times s_i are upper bounded by $\min(\pi, \pi/R)$, which is a constant which depends on the Hamiltonian H only. Second, to compute each individual time s_i, one simply needs to solve the problem

$$\frac{1-\cos s'}{1-\cos Rs'} = k$$

For some $k \in (\min(R^{-2}, R^2), \max(R^{-2}, R^2))$ and s' in $(0, \min(\pi, \pi/R))$. In the region of s where the value of this function is between $\min(R^{-2}, R^2)$ and $\max(R^{-2}, R^2))$, the derivatives of this function are bounded by a function of R only. Furthermore, the derivatives of these terms are computable to accuracy ε in time $\text{polylog}(1/\varepsilon)$ time using the Taylor series for sine and cosine. Hence Newton's method can be used to solve this problem, and will achieve quadratic convergence, i.e. for each step you run Newton's method, the error is squared, and the number of digits of accuracy achieved doubles. Hence one can compute each time t_i to accuracy ε in $\text{polylog}(1/\varepsilon)$ time as desired. Furthermore, since inverting any particular L only requires inverting some fixed $c \in \mathbb{R}^+$ using Claim 8, an error ε in an individual $N(s_i)$

matrices contributes $c\varepsilon$ error to the operator norm[6] of $N(s_1)...N(s_k)$, and hence $c\varepsilon$ error to the operator norm of L^{-1}. Hence this construction is "computationally efficient" for each fixed L as defined previously.

□

This completes the proof in Case 3 and hence the entire proof.

□

4.6 Showing density in $SL(2,\mathbb{C})$

We now prove Claim 5.

Proof of Claim 5. To show that $S = SL(2,\mathbb{C})$, we will first show that S is a group, and then show S is a Lie group.

Claim 9. *S is a group.*

Proof. Clearly, if we only took finite products of these elements, the resulting set of matrices would be a group, because we have the inverses of every element in the generating set. So what we need to show is that taking the closure of this set of matrices still yields a group. To see this, suppose that some element $s \in S \subseteq SL(2,\mathbb{C})$ is the limit of a sequence $L_1, L_2, \ldots$ where each L_i is a finite product of element of the form $L(D(t_1,t_2))$, and $\lim_{i\to\infty} L_i = s$. Now consider the sequence $L_1^{-1}, L_2^{-1}, \ldots$. We claim that $\lim_{i\to\infty} L_i^{-1} = s^{-1}$. To see this, simply note that for a 2×2 matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL(2,\mathbb{C})$, its inverse is given by $\left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$. Since the limit point s exists in $SL(2,\mathbb{C})$, the limit of each matrix entry of the L_i's must converge as well to the entries of s. Hence the entries of the sequence L_i^{-1} converges to the entries of s^{-1}. □

Note that it is critical that we've taken the closure in $SL(2,\mathbb{C})$; if we took the closure in the set of 2×2 complex matrices, this would not necessarily be true.

We have now established that S is a group. Furthermore, S is a closed subgroup of $SL(2,\mathbb{C})$ by construction, and $SL(2,\mathbb{C})$ is a Lie group. We now invoke a well-known theorem from Lie theory.

Theorem 4.6.1 (Cartan's Theorem [83] or the Closed Subgroup Theorem)**.** *Any closed subgroup of a Lie group is a Lie group.*

Corollary 4.6.2. *S is a Lie group.*

Now that we know S is a Lie group, we can use facts from Lie theory to show $S = SL(2,\mathbb{C})$. We provide a brief over of Lie theory in Section 2.3.3, but a more complete treatment can be found in e.g. [141] or a more advanced textbook on Lie groups.

To show that $S = SL(2,\mathbb{C})$, we will consider $\mathfrak{g} \triangleq \text{Lie}(S)$. We will then show that $\mathfrak{g} = \mathfrak{sl}(2,\mathbb{C})$, which is the Lie algebra of $SL(2,\mathbb{C})$, which consists of all traceless two by two complex matrices. Since the exponential map maps the algebra into the group (see Section 2.3.3), this implies that $\exp(\mathfrak{sl}(2,\mathbb{C})) \subseteq S$. From this, we will leverage the following fact:

Claim 10. $\exp(\mathfrak{sl}(2,\mathbb{C}))$ *is dense in* $SL(2,\mathbb{C})$.

[6]This is because for non-unitary matrices, the norm of the singular values are not one. Hence when considering the product AB, where λ_{max} is the largest singular value of A, an ε error in B will induce an $\lambda_{max}\varepsilon$ error in AB.

Proof. It is well known [141] that $\exp(\mathfrak{sl}(2,\mathbb{C}))$ contains all matrices in $SL(2,\mathbb{C})$ except matrices A for which $\mathrm{Tr}(A) = -2$ and $A \neq -I$. This implies the claim. □

Hence to prove Claim 5, it suffices to prove the following claim:

Claim 11. $\mathfrak{g} \triangleq \mathrm{Lie}(S)$ *spans* $\mathfrak{sl}(2,\mathbb{C})$, *i.e. all* 2×2 *traceless matrices.*

Proof. Consider elements of the form

$$M(t,s) \triangleq L(t)L(s)^{-1}.$$

As t, s vary over $(0,\pi) \cup (\pi, 2\pi)$, these form continuous paths within S. In particular, at the point where $s = t$, this path passes through the identity. Now consider

$$g(v) \triangleq \frac{\partial}{\partial t}\left[M(t,s)\right]\Big|_{s=t=v}$$

These are tangent vectors to paths in S, evaluated as they pass through the identity. Hence we have that $g(v) \in \mathfrak{g}$ for all $v \in (0,\pi) \cup (\pi, 2\pi)$. By direct calculation, one can show that

$$g(v) = -\frac{1}{2\sin(2v)} \begin{pmatrix} (a'+1)e^{-2iv} & \frac{\alpha}{\beta}(1-a')e^{i(1+a')v} \\ \frac{\beta}{\alpha}(3+a')e^{i(-1-a')v} & -(a'+1)e^{-2iv} \end{pmatrix}$$

where we have simplified using the fact that $d' = -2 - a'$.

We will now break into cases to show that these matrices span the entire Lie algebra. We begin with the generic case and then give the special cases. In the special cases, we will also add additional postselection gadgets to our model in order to get single-qubit transformations which span all traceless matrices. The gadgets introduced are inherently closed under taking inverses. So this simply reflects that for very particular Hamiltonians, our L matrices need additional help to span all 1-qubit operations.

Case 1: $a' \neq \pm 1, -3$

In this case all of the entries of $g(v)$ are non-zero.

$$g(v) = -\frac{1}{2\sin(2v)} \begin{pmatrix} (a'+1)e^{-2iv} & \frac{\alpha}{\beta}(1-a')e^{i(1+a')v} \\ \frac{\beta}{\alpha}(3+a')e^{i(-1-a')v} & -(a'+1)e^{-2iv} \end{pmatrix}$$

We can therefore rewrite $g(v)$ with four non-zero parameters $k_1 \in \mathbb{R}$, $k_2, k_3 \in \mathbb{C}$, and using a new parameter $v' = -2v$:

$$g(v) \propto \begin{pmatrix} e^{v'} & k_2 e^{ik_1 v'} \\ k_3 e^{-ik_1 v'} & -e^{iv'} \end{pmatrix}$$

Here we omit real coefficients as the Lie algebra is closed under scalar multiplication by $\mathbb{R}$. The fact that $a' \neq \pm 1, -3$ also implies that $k_4 \neq \pm 1$

Now consider the value of $g(v')$ for small values of v'. In particular, pick a $\theta << 1$. Then we have that

$$g(\pm\theta) \propto \begin{pmatrix} (A \pm Bi) & k_2(C \pm Di) \\ k_3(C \mp Di) & -(A \pm Bi) \end{pmatrix}$$

for some nonzero real coefficients $A, B, C, D \in \mathbb{R}$. Taking the sum and difference of these matrices, we see the following are elements of the Lie algebra:

$$\begin{pmatrix} A & k_2 C \\ k_3 C & -A \end{pmatrix} \quad \begin{pmatrix} Bi & k_2 Di \\ -k_3 Di & -Bi \end{pmatrix}$$

Likewise, by considering taking the sum and difference of $g(\pm 2\theta)$, we get there exist nonzero $A', B', C', D' \in \mathbb{R}$ such that the lie algebra contains.

$$\begin{pmatrix} A' & k_2 C' \\ k_3 C' & -A' \end{pmatrix} \quad \begin{pmatrix} B'i & k_2 D'i \\ -k_3 D'i & -B'i \end{pmatrix}$$

Furthermore, since sine and cosine are nonlinear, and $k_1 \neq 0, \pm 1$, the vectors (A, C) and (A', C') are linearly independent. Likewise the vectors (B, D) and (B', D') are linearly independent. Hence by taking linear combinations of these matrices, we have that any matrix of the form

$$\begin{pmatrix} E & k_2 F \\ k_3 F^* & -E \end{pmatrix}$$

is in the Lie algebra for any $E, F \in \mathbb{C}$. Hence our Lie algebra spans at least these two complex dimensions. Now we take the closure of such matrices under commutators. Suppose $A, B, C, D \in \mathbb{C}$. We have that

$$\left[\begin{pmatrix} A & k_2 B \\ k_3 B^* & -A \end{pmatrix}, \begin{pmatrix} C & k_2 D \\ k_3 D^* & C \end{pmatrix} \right] = \begin{pmatrix} k_2 k_3 (BD^* - B^* D) & 2k_2 (AD - BC) \\ 2k_3 (B^* C - AD^*) & k_2 k_3 (B^* D - BD^*) \end{pmatrix}$$

Since we previously showed all traceless diagonal matrices are in the Lie algebra, this implies the following matrices are in the Lie algebra:

$$\begin{pmatrix} 0 & 2k_2 (AD - BC) \\ 2k_3 (B^* C - AD^*) & 0 \end{pmatrix}$$

By setting A, D, B, C such that $(AD - BC)^* \neq (B^* C - AD^*)$, we can see that these matrices span the remaining two real dimensional space of off-diagonal matrices. Hence our Lie algebra spans all traceless matrices. This completes the proof in Case 1.

Case 2: $a' = 1$ or $a' = -3$

We will prove the claim for $a' = 1$; an analogous proof holds for $a' = -3$. (These are the Hamiltonians $\mathrm{diag}(1, 1, 1, -3)$ and $\mathrm{diag}(-3, 1, 1, 1)$, which are identical except the role of 0 and 1 is switched.).

In this case we have that

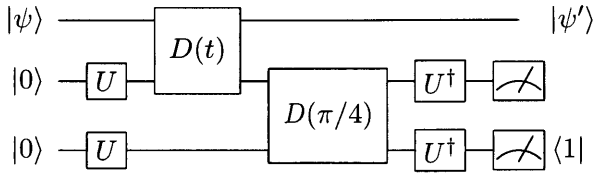
$$g(v) \propto \begin{pmatrix} e^{-2iv} & 0 \\ 2\dfrac{\beta}{\alpha} e^{-2iv} & -e^{-2iv} \end{pmatrix}$$

By evaluating $g(v)$ at $\pm\theta$ and $\pm 2\theta$ for some small value of θ, by the same arguments put forth in Case 1, these matrices span the space of matrices of the form

$$\begin{pmatrix} A + Bi & 0 \\ 2\dfrac{\beta}{\alpha}(A + Bi) & -A - Bi \end{pmatrix}$$

Where $A, B \in \mathbb{R}$ are arbitrary real parameters.

We will now use another postselection gadget, which is inherently closed under taking inverses, to boost the span of the algebra to all of $\mathfrak{sl}(2, \mathbb{C})$. This is the same gadget which appears in the construction of L^{-1} in appendix 4.5.



This gadget performs the operation

$$P(t) \propto \begin{pmatrix} e^{it} & 0 \\ 0 & e^{-3it} \end{pmatrix} \propto \begin{pmatrix} e^{2it} & 0 \\ 0 & e^{-2it} \end{pmatrix}$$

Hence its Lie algebra spans the space of traceless diagonal imaginary matrices. Combining this with the previous result, we see the Lie algebra now spans the space

$$\begin{pmatrix} A+Bi & 0 \\ 2\frac{\beta}{\alpha}(A+Ci) & -A-Bi \end{pmatrix}$$

Where $A, B, C \in \mathbb{R}$ are arbitrary real parameters.

Now consider taking commutators of such matrices; one can easily see that for $A, B, C, D, E, F \in \mathbb{R}$,

$$\left[\begin{pmatrix} A+Bi & 0 \\ 2\frac{\beta}{\alpha}(A+Ci) & -A-Bi \end{pmatrix}, \begin{pmatrix} D+Ei & 0 \\ 2\frac{\beta}{\alpha}(D+Fi) & -D-Ei \end{pmatrix}\right] = \begin{pmatrix} 0 & 0 \\ 4\frac{\beta}{\alpha}(A+Ci)(D+Ei) & 0 \end{pmatrix}$$

Hence by appropriate choice of A, C, D, E these commutators span all complex values in the lower left hand corner. So our Lie algebra now spans

$$\begin{pmatrix} A+Bi & 0 \\ C+Di & -A-Bi \end{pmatrix}$$

Where $A, B, C, D \in \mathbb{R}$ are arbitrary real parameters. In other words we span all traceless lower triangular matrices.

Next we will use the fact that the Lie algebra is closed under conjugation by the group. Therefore it must contain all elements of the form

$$L(t) \begin{pmatrix} A & 0 \\ B & -A \end{pmatrix} L^{-1}(t)$$

where A, B are now complex parameters

Since we already span lower triangular matrices, the only relevant entry of the above matrix is the upper-right entry, as we can zero out the other entries by adding lower triangular matrices. This upper left entry is proportional to

$$i\left(-2\alpha\beta^*|\alpha|^2 e^{2it} A - \alpha^2 \beta^{*2} e^{2it} B\right)$$

Since α and β are non-zero, and setting $B = 0$, we can see that by choosing A we can set this value to be any complex number. Hence our Lie algebra must span

$$L(t)\begin{pmatrix} A & C \\ B & -A \end{pmatrix} L^{-1}(t)$$

Where $A, B, C \in \mathbb{C}$, that is all of $\mathfrak{sl}(2,\mathbb{C})$, as desired. This completes the proof of Claim 2.

Case 3: $a' = -1$

In this case we have that

$$g(v) = -\frac{1}{\sin(2v)}\begin{pmatrix} 0 & \frac{\alpha}{\beta} \\ \frac{\beta}{\alpha} & 0 \end{pmatrix}$$

Thus the matrices $g(v)$ span a one-dimensional space. Since the Lie algebra is closed under scalar multiplication by reals, the factor of $\frac{-1}{\sin(2v)}$ out front is irrelevant, and we will drop real prefactors in future calculations.

We will now use the fact the Lie algebra is closed under conjugation by the group. Consider matrices of the form

$$T(s,v) = L(s)g(v)L(s)^{-1} \propto i\begin{pmatrix} |\beta|^4 - |\alpha|^4 & |\alpha|^4\frac{\alpha}{\beta}e^{-2is} - \alpha\beta^*|\beta|^2e^{2is} \\ \frac{\beta}{\alpha}|\beta|^4e^{-2is} - \alpha^*\beta|\alpha|^2e^{2is} & |\alpha|^4 - |\beta|^4 \end{pmatrix}$$

where the proportionality is over real scalar multiples. Here we have simplified using the fact we are in the case $a' = d' = -1$. This is well defined for any s and v which are not integer multiples of π.

Now we break into two subcases:

Subcase A: $|\alpha|^2 \neq |\beta|^2$

In this case, the matrix $T(s,v)$ has a nonzero entry on the diagonals. Hence the matrix $T(s,v)$ has the form

$$T(s,v) \propto i\begin{pmatrix} k_1 & k_2e^{-2is} - k_3e^{2is} \\ k_4e^{-2is} - k_5e^{2is} & -k_1 \end{pmatrix}$$

Where $k_1 \in \mathbb{R}$ is nonzero, $k_2, k_3, k_4, k_5 \in \mathbb{C}$ are nonzero. One can easily check that the constraint $|\alpha|^2 \neq |\beta|^2$ further implies that k_2, k_3, k_4, k_5 have four distinct values, i.e. $k_i \neq k_j$ for any $i \neq j$, $i, j \geq 2$. For instance, to see that $k_2 \neq k_3$, note that if $k_2 = k_3$ then $|\alpha|^4\frac{\alpha}{\beta} = \alpha\beta^*|\beta|^2$, which implies $|\alpha|^4 = |\beta|^4$, a contradiction.

Furthermore, one can show that there cannot exist a constant[7] K such that $k_2 = Kk_4$ and $k_3 = Kk_5$, because this would imply $|K| = |\frac{\alpha}{\beta}|^6 = |\frac{\alpha}{\beta}|^2$ which is a contradiction if

[7]If this were the case, the matrices $T(s,v)$ would only span matrices of the form $\begin{pmatrix} Ai & B+Ci \\ K(B+Ci) & -Ai \end{pmatrix}$. Fortunately this does not happen in this case.

$|\alpha| \neq |\beta|$. Hence the matrices $T(s,v)$ span matrices of the form

$$\begin{pmatrix} Ai & B+Ci \\ D+Ei & -Ai \end{pmatrix}$$

where $A, B, C, D, E \in \mathbb{R}$ are arbitrary real parameters. Now taking the closure of such matrices under commutators, one can easily see this spans all traceless matrices. Hence the Lie algebra spans $\mathfrak{sl}(2,\mathbb{C})$ as desired.

Subcase B: $|\alpha|^2 = |\beta|^2 = 1/2$

In this case the Hamiltonians generated are of the form $X(\theta) \otimes X(\theta)$, so are not covered in the scope of this theorem. Note that the Lie algebra of the L gadgets here only span a two dimensional subspace of the form

$$\begin{pmatrix} 0 & e^{-i\theta}(A+Bi) \\ e^{i\theta}(A+Bi) & 0 \end{pmatrix}$$

where $A, B \in \mathbb{R}$. This is closed under conjugation and does not span $\mathfrak{sl}(2,\mathbb{C})$.

□

□

4.7 Proof of postselected universality when $b \neq c$

Here we consider the postselected universality of circuits with entangling Hamiltonians for which $H \neq THT$. The proof in this case will follow analogously to the main proof. Furthermore, the construction of the inverse gadgets will have a much cleaner construction than the case $H = THT$.

Suppose we have a commuting Hamiltonian H such that $H \neq THT$. By Lemma 4.1.3, we know that $H = (U \otimes U)\,\mathrm{diag}(a,b,c,d)(U^\dagger \otimes U^\dagger)$ for some one-qubit unitary $U = \left(\begin{smallmatrix} \alpha & -\beta^* \\ \beta & \alpha^* \end{smallmatrix}\right)$ and some real parameters a, b, c, d. The trace of H contributes an irrelevant global phase to the unitary operator it generates, so without loss of generality we can assume H is traceless, i.e., $a+b+c+d = 0$. Since $H \neq THT$ we have $b \neq c$. As before, the fact H can generate entanglement starting from the computational basis implies $\alpha \neq 0, \beta \neq 0$, and $b + c \neq 0$.

Now consider the Hamiltonians

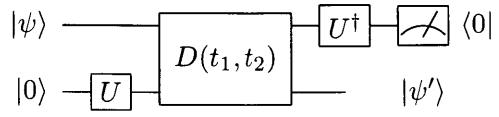
$$H_1 = \frac{1}{c^2 - b^2}(cH_{12} - bH_{21}), \quad H_2 = \frac{1}{b^2 - c^2}(bH_{12} - cH_{21})$$

Since we can apply both H, $-H$, THT, and $-THT$, this allows us to apply H_1 and H_2 for independent amounts of time. Let $V(t_1, t_2)$ be the two-qubit unitary we obtain from running H_1 for time $t_1 \in \mathbb{R}$ and H_2 for time $t_2 \in \mathbb{R}$. We have

$$V(t_1,t_2) = e^{it_1 H_1} e^{it_2 H_2} = (U^{\otimes 2}) D(t_1,t_2) (U^{\dagger \otimes 2}),$$

where $D(t_1,t_2) \triangleq \mathrm{diag}(e^{ia'(t_1+t_2)}, e^{it_1}, e^{it_2}, e^{id'(t_1+t_2)})$.

Now following our previous proof, we consider the following postselection gadget:



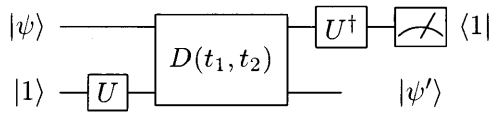
This performs the following transformation on the input state:

$$L(t_1,t_2) = \frac{1}{|\alpha||\beta|\sqrt{\left(e^{-i(t_1+t_2)} - e^{i(t_1+t_2)}\right)}} \begin{pmatrix} |\alpha|^2 e^{ia'(t_1+t_2)} & \alpha\beta^* e^{it_2} \\ \alpha^*\beta e^{it_1} & |\beta|^2 e^{id'(t_1+t_2)} \end{pmatrix}.$$

As before, this is a non-unitary transformation, and hence it is unclear how to invert L. Fortunately, when $H \neq THT$ we have the freedom to apply H_1 and H_2 for separate times, and this allows us to make a much simpler postselecting gadget to invert L, as follows:

Claim 12. *Given $L(t_1,t_2)$, where $t_i \in (0,\pi)\cup(\pi,2\pi)$, it is possible to construct $L(t_1,t_2)^{-1}$ by introducing three postselections into the circuit. Furthermore, this construction is efficiently computable in the manner described above.*

Proof. We will need two additional gadgets for our construction. First, consider a modification of the gadget for $L(t_1,t_2)$, where we start the qubit in the $|1\rangle$ state and postselect on the $|1\rangle$ state:



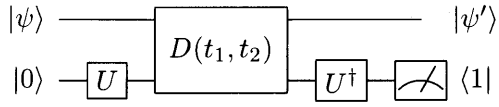
By a direct calculation, one can show the linear transformation performed on $|\psi\rangle$ is given by

$$M(t_1,t_2) = \frac{1}{|\alpha||\beta|\sqrt{\left(e^{-i(t_1+t_2)} - e^{i(t_1+t_2)}\right)}} \begin{pmatrix} |\beta|^2 e^{ia'(t_1+t_2)} & -\alpha\beta^* e^{it_2} \\ -\alpha^*\beta e^{it_1} & |\alpha|^2 e^{id'(t_1+t_2)} \end{pmatrix}$$

This is tantalizingly close to the inverse of L, which is

$$L(t_1,t_2)^{-1} = \frac{1}{|\alpha||\beta|\sqrt{\left(e^{-i(t_1+t_2)} - e^{i(t_1+t_2)}\right)}} \begin{pmatrix} |\beta|^2 e^{id'(t_1+t_2)} & -\alpha\beta^* e^{it_2} \\ -\alpha^*\beta e^{it_1} & |\alpha|^2 e^{ia'(t_1+t_2)} \end{pmatrix}$$

The only thing that is off is that the phase of the upper left and bottom right entries are incorrect. To correct these phases, we need to introduce another gadget:



In other words, instead of using the gate in a teleportation-like protocol, we instead use it to apply phases to $|\psi'\rangle$. This gate performs the following transformation on the input state:

$$N(t_1,t_2) = \frac{1}{\sqrt{(e^{it_1} - e^{ia'(t_1+t_2)})(e^{id'(t_1+t_2)} - e^{it_2})}} \begin{pmatrix} e^{it_1} - e^{ia'(t_1+t_2)} & 0 \\ 0 & e^{id'(t_1+t_2)} - e^{it_2} \end{pmatrix}$$

Since N is a diagonal matrix, the only physical quantity that matters is the ratio $r(t_1,t_2)$ of its two entries, which is a complex number given by

$$r(t_1,t_2) = \frac{e^{it_1} - e^{ia'(t_1+t_2)}}{e^{id'(t_1+t_2)} - e^{it_2}}.$$

If $r = r(t_1,t_2)$ takes on a certain value, then it immediately follows that $N(t_1,t_2) =$

$\begin{pmatrix} \sqrt{r} & 0 \\ 0 & \sqrt{r}^{-1} \end{pmatrix}$, because of our normalization.

We will now show that by setting t_1 and t_2, we can choose $r(t_1, t_2)$ to be any complex phase $e^{i\theta}$ that we like. In fact, if $\frac{a'}{d'}$ is irrational, one can also show that one can choose t_1, t_2 to approximate any complex number; however, this will not be necessary for our construction, so we omit this here.

Claim 13. *For any $\theta \in (0, 2\pi)$, there exist $t_1, t_2 \in \mathbb{R}$ such that $r(t_1, t_2) = e^{i\theta}$.*

Proof. Set $t_1 = \theta$ and $t_2 = -\theta$. We immediately have

$$r(\theta, -\theta) = \frac{e^{i\theta} - 1}{1 - e^{-i\theta}} = \frac{e^{i\theta} - 1}{e^{-i\theta}(e^{i\theta} - 1)} = e^{i\theta}.$$

Note that this only works if $e^{i\theta} \neq 1$ - this is why we have omitted $\theta = 0$ from our range of θ. In other words, this gadget can be used to perform any diagonal matrix other than the identity. □

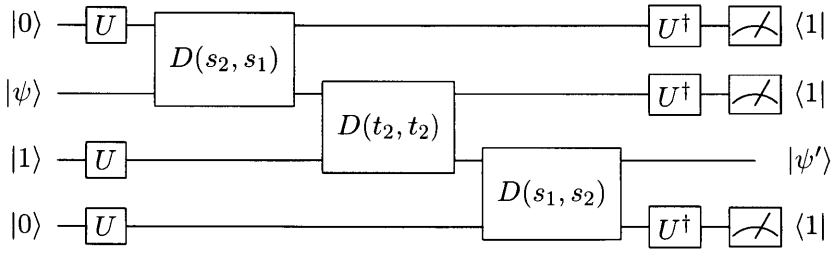
Putting this all together, we now show how to invert $L(t_1, t_2)$. Set $s_1 = i(d'(t_1 + t_2) - a'(t_1 + t_2))$ and $s_2 = -s_1$. Then we have[8]

$$N(s_1, s_2) = \begin{pmatrix} e^{\frac{i}{2}(d'(t_1+t_2)-a'(t_1+t_2))} & 0 \\ 0 & e^{-\frac{i}{2}(d'(t_1+t_2)-a'(t_1+t_2))} \end{pmatrix}$$

Now one can easily check that

$$L(t_1, t_2)^{-1} = N(s_1, s_2) M(t_1, t_2) N(s_1, s_2)$$

And therefore the following gadget performs $L(t_1, t_2)^{-1}$:



(Note that s_1 and s_2 are switched in the first diagonal matrix, as we have switched the usual order of the qubits.)

Hence using these postselection gadgets, we can generate not only $L(t_1, t_2)$, but also its inverse. Furthermore, this construction is manifestly efficient, since s_1 and s_2 are efficiently computable given t_1 and t_2. □

We can therefore apply both $L(t_1, t_2)$ and $L(t_1, t_2)^{-1}$ in our postselected circuits. This once again allows us to apply Lie theory to determine which subset of transformations can be applied by taking products of L matrices. Following our proof of the main theorem, we

[8] This is possible as long as $e^{i(d'(t_1+t_2)-a'(t_1+t_2))} \neq 1$. If this quantity is one, then $L(t_1, t_2)^{-1} = M(t_1, t_2)$, so no additional gadgets are necessary to obtain inverses.

now show the Lie algebra of the L matrices spans $\mathfrak{sl}(2,\mathbb{C})$. This completes the proof of postselected universality in this case in analogy with the main theorem.

Claim 14. *The Lie algebra of the L matrices spans $\mathfrak{sl}(2,\mathbb{C})$ in the case where $T \neq THT$.*

Proof. Consider elements of the form

$$M(t_1,t_2,s_1,s_2) \triangleq L(D(t_1,t_2))L(D(s_1,s_2))^{-1}.$$

As t_1,t_2,s_1,s_2 vary over the set

$$\{t_1,t_2 : t_1+t_2 \in (0,\pi)\cup(\pi,2\pi)\} \times \{s_1,s_2 : s_1+s_2 \in (0,\pi)\cup(\pi,2\pi)\}$$

, these form continuous paths within S. In particular, at the point where $s_1 = t_1$ and $s_2 = t_2$, this path passes through the identity. Now consider

$$g(v_1,v_2) \triangleq \frac{\partial}{\partial t_1}\left[M(t_1,t_2,s_1,s_2)\right]\Big|_{\substack{s_1=t_1=v_1\\ s_2=t_2=v_2}}$$

and

$$h(v_1,v_2) \triangleq \frac{\partial}{\partial t_2}\left[M(t_1,t_2,s_1,s_2)\right]\Big|_{\substack{s_1=t_1=v_1\\ s_2=t_2=v_2}}.$$

These are tangent vectors to paths in S, evaluated as they pass through the identity. Hence we have that $g(v_1,v_2)$ and $h(v_1,v_2) \in \mathfrak{g}$ for all $v_1,v_2 \in \{v_1,v_2 : v_1+v_2 \in (0,\pi)\cup(\pi,2\pi)\}$. By direct calculation, one can show that

$$g(v_1,v_2) = -\frac{1}{2\sin(v_1+v_2)}\begin{pmatrix} a'e^{-i(v_1+v_2)}+\cos(v_1+v_2) & -\frac{\alpha}{\beta}a'e^{i(a'v_1+(a'+1)v_2)} \\ \frac{\beta}{\alpha}(2+a')e^{i((d'+1)v_1+d'v_2)} & -a'e^{-i(v_1+v_2)}-\cos(v_1+v_2)\end{pmatrix}$$

and

$$h(v_1,v_2) = -\frac{1}{2\sin(v_1+v_2)}\begin{pmatrix} a'e^{-i(v_1+v_2)}-i\sin(v_1+v_2) & \frac{\alpha}{\beta}(1-a')e^{i(a'v_1)+(a'+1)v_2} \\ \frac{\beta}{\alpha}(1+a')e^{i((d'+1)v_1+d'v_2)} & -a'e^{-i(v_1+v_2)}+i\sin(t+1+v_2)\end{pmatrix}$$

where we have simplified using the fact that $d' = -1-a'$. Now suppose that we evaluate these matrices at the points where $v_1 = \theta$ and $v_2 = \frac{\pi}{2}-\theta$ for some real parameter θ; this ensures that v_1,v_2 are in the allowed set, and simplifies the above expressions to

$$\begin{aligned} g(\theta) &= -\frac{1}{2}\begin{pmatrix} -a'i & -\frac{\alpha}{\beta}a'e^{i(-\theta+(a'+1)\frac{\pi}{2})} \\ \frac{\beta}{\alpha}(2+a')e^{i(\theta+d'\frac{\pi}{2})} & a'i\end{pmatrix} \\ &= -\frac{1}{2}\begin{pmatrix} -a'i & -\frac{\alpha}{\beta}a'e^{i\theta'} \\ \frac{\beta}{\alpha}(2+a')e^{-i\theta'} & a'i\end{pmatrix}, \end{aligned}$$

here we define $\theta' = -\theta+(a'+1)\frac{\pi}{2}$; this follows from the fact that $d' = -1-a'$. Likewise,

we can consider $h(v_1, v_2)$ evaluated when $v_1 = \theta$ and $v_2 = \frac{\pi}{2} - \theta$; this evaluates to

$$h(\theta) = -\frac{1}{2}\begin{pmatrix} -ia' - i & \frac{\alpha}{\beta}(1-a')e^{i(-\theta+(a'+1)\frac{\pi}{2})} \\ \frac{\beta}{\alpha}(1+a')e^{i(\theta+a'\frac{\pi}{2})} & ia'+i \end{pmatrix}$$
$$= -\frac{1}{2}\begin{pmatrix} -i(a'+1) & \frac{\alpha}{\beta}(1-a')e^{i\theta'} \\ \frac{\beta}{\alpha}(1+a')e^{-i\theta'} & i(a'+1) \end{pmatrix}.$$

By setting the value of θ in the range $[0, 2\pi)$, we can select any values of θ' we like; hence we will work with θ' from this point forward.

For now we will assume that $a' \neq 0$ and $a' \neq 1$; we will handle the cases $a' = 0$ and $a' = -1$ separately. The proof of the general case is the most difficult one.

Case 1: $a' \neq 0$ and $a' \neq -1$.

We know that $g(\theta') \in \mathfrak{g}$ and $h(\theta') \in \mathfrak{g}$. Furthermore, since $\mathfrak{g}$ is a real Lie algebra, it is closed as a vector space over $\mathbb{R}$. Hence we must also have that

$$j(\theta_1, \theta_2) \triangleq -2\left(\frac{1}{a'+1}h(\theta_2) - \frac{1}{a'}g(\theta_1)\right) = \begin{pmatrix} 0 & \frac{\alpha}{\beta}\left(\frac{1-a'}{1+a'}e^{i\theta_2} + e^{i\theta_1}\right) \\ \frac{\beta}{\alpha}\left(e^{-i\theta_2} - \frac{2+a'}{a'}e^{-i\theta_1}\right) & 0 \end{pmatrix} \in \mathfrak{g}$$

Where we have used the assumption that $a' \neq 0$ and $a' \neq -1$. We will now show that as we vary θ_1 and θ_2, these elements $j(\theta_1, \theta_2)$ span all two by two matrices of the form $\left(\begin{smallmatrix} 0 & c_1 \\ c_2 & 0 \end{smallmatrix}\right)$, where $c_1, c_2 \in \mathbb{C}$.

To prove this, we will break into two subcases. For convenience, define

$$k = \frac{a'-1}{a'+1}.$$

Subcase A: $a' > 0$, i.e., $-1 < k < 1$.

In this subcase, consider the matrices

$$\frac{-a'(1+a')}{4}\left[j\left(\arcsin k, \frac{\pi}{2}\right) + j\left(\pi - \arcsin k, \frac{\pi}{2}\right)\right] = \begin{pmatrix} 0 & 0 \\ \frac{\beta}{\alpha}i & 0 \end{pmatrix} \tag{4.11}$$

$$\frac{1+a'}{4\sqrt{a'}}\left[j\left(\arcsin k, \frac{\pi}{2}\right) - j\left(\pi - \arcsin k, \frac{\pi}{2}\right)\right] = \begin{pmatrix} 0 & \frac{\alpha}{\beta} \\ -\frac{\beta}{\alpha}\frac{2+a'}{a'} & 0 \end{pmatrix} \tag{4.12}$$

and

$$\frac{a'(1+a')}{4}\left[j\left(\arccos k, 0\right) + j\left(-\arccos k, 0\right)\right] = \begin{pmatrix} 0 & 0 \\ \frac{\beta}{\alpha} & 0 \end{pmatrix} \tag{4.13}$$

$$\frac{1+a'}{4\sqrt{a'}}\left[j\left(\arccos k, 0\right) - j\left(-\arccos k, 0\right)\right] = \begin{pmatrix} 0 & \frac{\alpha}{\beta}i \\ \frac{\beta}{\alpha}\frac{2+a'}{a'}i & 0 \end{pmatrix}. \tag{4.14}$$

These are well-defined as we have $a' > 0$ in this case. Clearly matrices (4.11) and (4.13) span the space of all matrices with a single complex entry in the bottom left hand corner. Hence, when combined with matrices (4.12) and (4.14), they clearly span the space of all matrices with complex entries in the off diagonal elements.

Subcase B: $a' < 0$ and $a' \neq -1$, i.e., $-1 < 1/k < 1$

This subcase follows similarly; consider the matrices

$$\frac{a'(1-a')}{4}\left[j\left(\frac{\pi}{2}, \arcsin\frac{1}{k}\right) + j\left(\frac{\pi}{2}, \pi - \arcsin\frac{1}{k}\right)\right] = \begin{pmatrix} 0 & 0 \\ \frac{\beta}{\alpha}i & 0 \end{pmatrix} \tag{4.15}$$

$$\frac{1+a'}{4\sqrt{-a'}}\left[j\left(\frac{\pi}{2}, \arcsin\frac{1}{k}\right) - j\left(\frac{\pi}{2}, \pi - \arcsin\frac{1}{k}\right)\right] = \begin{pmatrix} 0 & \frac{\alpha}{\beta} \\ -\frac{\beta}{\alpha}\frac{1+a'}{1-a'} & 0 \end{pmatrix} \tag{4.16}$$

and

$$\frac{-a'(1-a')}{4}\left[j\left(0, \arccos\frac{1}{k}\right) + j\left(0, -\arccos\frac{1}{k}\right)\right] = \begin{pmatrix} 0 & 0 \\ \frac{\beta}{\alpha} & 0 \end{pmatrix} \tag{4.17}$$

$$\frac{1+a'}{4\sqrt{-a'}}\left[j\left(0, \arccos\frac{1}{k}\right) - j\left(0, -\arccos\frac{1}{k}\right)\right] = \begin{pmatrix} 0 & \frac{\alpha}{\beta}i \\ \frac{\beta}{\alpha}\frac{1+a'}{1-a'}i & 0 \end{pmatrix}. \tag{4.18}$$

These are well-defined as we have $a' < 0$ in this case, as well as $a' \neq -1$. Again, clearly we have that (4.15) and (4.17) span all matrices with a single complex entry in the bottom left of the matrix. Hence, adding in (4.16) and (4.18), we span all off-diagonal complex matrices, which is what we wanted to show.

In either subcase, our j matrices span all matrices of the form

$$\begin{pmatrix} 0 & A+Bi \\ C+Di & 0 \end{pmatrix}$$

where $A, B, C, D \in \mathbb{R}$. Additionally, our g and h matrices are also in $\mathfrak{g}$, and clearly combining these with the j matrices increases the span to

$$\begin{pmatrix} Ei & A+Bi \\ C+Di & -Ei \end{pmatrix}$$

where $A, B, C, D, E \in \mathbb{R}$. This is a five-dimensional subspace of $\mathfrak{sl}(2, \mathbb{C})$. Now to show that we can span all 6 dimensions of $\mathfrak{sl}(2, \mathbb{C})$, we invoke the fact that $\mathfrak{g}$ is closed under commutation, so $\mathfrak{g}$ contains $[\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)] = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Hence $\mathfrak{g}$ must include all matrices of the form

$$\begin{pmatrix} F+Ei & A+Bi \\ C+Di & -F-Ei \end{pmatrix}$$

where $A, B, C, D, E, F \in \mathbb{R}$. In other words, $\mathfrak{g} = \mathfrak{sl}(2, \mathbb{C})$.

We've now shown Claim 11 in the case where $a' \neq 0$ and $a' \neq -1$. We now prove the claim in these remaining two cases.

Case 2: $a' = 0$.

In this case we have

$$g(\theta') = -\frac{1}{2}\begin{pmatrix} 0 & 0 \\ \frac{\beta}{\alpha}2e^{-i\theta'} & 0 \end{pmatrix}$$

As θ varies these matrices clearly span all matrices a single complex number in the bottom left entry. Now in this case we also have that

$$h(\theta') = -\frac{1}{2}\begin{pmatrix} -i & \frac{\alpha}{\beta}e^{i\theta'} \\ \frac{\beta}{\alpha}e^{-i\theta'} & i \end{pmatrix}$$

Since $\mathfrak{g}$ is closed under addition and scalar multiplication by $\mathbb{R}$, and applying

$$h(\theta') - h(\theta'') = -\frac{1}{2}\begin{pmatrix} 0 & \frac{\alpha}{\beta}(e^{i\theta'} - e^{i\theta''}) \\ \frac{\beta}{\alpha}(e^{-i\theta'} - e^{i\theta''}) & 0 \end{pmatrix} \in \mathfrak{g}$$

Now adding in multiples of g, we have that $\mathfrak{g}$ contains matrices of the form

$$\begin{pmatrix} 0 & \frac{\alpha}{\beta}(e^{i\theta'} - e^{i\theta''}) \\ 0 & 0 \end{pmatrix}$$

which clearly span all matrices with a complex entry in the upper right corner. Hence we span all off-diagonal matrices. Now adding in $h(\theta)$ for any θ, we span all matrices of the form $\begin{pmatrix} Ei & A+Bi \\ C+Di & -Ei \end{pmatrix}$ where $A, B, C, D, E \in \mathbb{R}$. As discussed in Case 1, by taking the closure of these under commutation we have that $\mathfrak{g} = \mathfrak{sl}(2\mathbb{C})$ as desired, which completes the proof of Case 2.

Case 3: $a' = -1$

This case follows very similarly to Case 2. When $a' = -1$ we have that

$$h(\theta') = -\frac{1}{2}\begin{pmatrix} 0 & \frac{\alpha}{\beta}2e^{i\theta'} \\ 0 & 0 \end{pmatrix}$$

which clearly span all complex matrices with a single entry in the upper right corner. In this case, we also have that

$$g(\theta') = -\frac{1}{2}\begin{pmatrix} i & -\frac{\alpha}{\beta} - e^{i\theta'} \\ \frac{\beta}{\alpha}e^{-i\theta'} & -i \end{pmatrix},$$

By considering the difference $g(\theta') - g(\theta'')$, and noting that we already span matrices with a single entry in the upper right corner, this shows that we span all off-diagonal matrices. Now adding in $g(\theta')$ for any θ' we see that we span all matrices of the form $\begin{pmatrix} Ei & A+Bi \\ C+Di & -Ei \end{pmatrix}$ where $A, B, C, D, E \in \mathbb{R}$. As discussed in Case 1, by taking the closure of these under commutation we have that $\mathfrak{g} = \mathfrak{sl}(2, \mathbb{C})$ as desired. This completes the proof of Case 3, hence the proof of the claim. $\square$

4.8 Open Problems

Our results leave a number of open problems.

1. An interesting open problem is to classify *all* Hamiltonians in terms of their computational power under this model. Childs et al. [90] previously classified which two-qubit Hamiltonians can perform any unitary on two qubits. However, this does not classify which Hamiltonians are computationally universal for two reasons. First, as Childs et al. point out in their paper, it is possible that H fails to generate all unitaries on two qubits, but does generate all unitaries on three qubits (i.e. adding ancillae helps one attain universality). It remains open to classify which two-qubit H generate all unitaries on sufficiently large systems. Second, even if a Hamiltonian H does not generate all unitaries, it is still possible that H is computationally universal. For example, H could be universal on an encoded subspace. Classifying which Hamiltonians are universal under an encoding seems to be a challenging task. We conjecture that the power of any two-qubit Hamiltonian obeys a dichotomy: either H is efficiently classically simulable in this model, or it is universal under postselection and hence cannot be weakly simulated unless PH collapses. This is true of all known two-qubit Hamiltonians, and our classification proves this result rigorously in the case of commuting Hamiltonians.

2. In this chapter we considered the power of quantum circuits with commuting Hamiltonians. A more difficult related problem is classify the power of quantum circuits with commuting gate sets. The challenge in solving this problem would be to classify when a discrete set of L's generates a continuum of gates. There are some sufficient conditions under which this holds (see e.g. Aharonov et al. [29], Corollary 9.1). However, finding necessary and sufficient conditions under which a finite set of operators densely generates a continuous subgroup of $SL(2,\mathbb{C})$ seems very difficult, in part because there is no complete, explicit classification of discrete subgroups of $SL(2,\mathbb{C})$. Indeed, discrete subgroups of $SL(2,\mathbb{C})$ are related to the theory of Möbius transformations [53], where they are known as "Kleinian subgroups," and they are the subject of a deep area of mathematical research.

Chapter 5

Conjugated Clifford Circuits

In this chapter, we introduce a new model of "weak" quantum computation, which we call Conjugated Clifford Circuits (CCCs). The Gottesman-Knill Theorem [133] states that Clifford circuits - i.e. circuits composed of only CNOT, Hadamard, and $\pi/4$ phase gates - are efficiently classically simulable. We show that in contrast, "Conjugated Clifford Circuits" - where one additionally conjugates every qubit by the same one-qubit gate U - can perform hard sampling tasks. In particular, we fully classify the computational power of CCCs by showing that any non-Clifford conjugating unitary U can give rise to sampling tasks which cannot be exactly simulated classically, unless the polynomial hierarchy collapses. Furthermore, we show this hardness result can be extended to more realistic constant additive error under a plausible complexity-theoretic conjecture.

This chapter is based on joint work with Joseph Fitzsimons and Dax Koh [65].

5.1 Introduction

5.1.1 Our results

This chapter considers a new "weak" model of quantum computation which we call Conjugated Clifford Circuits (CCCs). (For a review of weak models of quantum computing, see section 2.2). In this model, we consider the power of quantum circuits which begin in the state $|0\rangle^{\otimes n}$, and then apply gates of the form $(U \otimes U)g(U^\dagger \otimes U^\dagger)$ where U is a fixed one-qubit gate and g is a Clifford gate (i.e. a gate from the set CNOT, H, P). In other words, we consider the power of Clifford circuits which are conjugated by an identical one-qubit gate U on each qubit. These gates manifestly perform a discrete subset of unitaries[1].

If U is the identity, then clearly this model is efficiently classically simulable by the Gottesman-Knill Theorem [133]. However, the presence of generic conjugating unitaries (even the same U on each qubit, as in this model) breaks this simulation algorithm[2]. This, combined with prior results showing hardness for other modified versions of Clifford circuits [163, 173], leads one to suspect that CCCs may not be efficiently classically simulable.

In this work, we confirm this intuition and provide two results in this direction. First, we show that CCCs cannot be efficiently classically simulated exactly (or to multiplicative

[1]Note that Bremner Jozsa and Shepherd's IQP hardness construction uses a discrete gate set - namely Controlled-Z and T, conjugated by Hadamards, or CZ,CCZ, Z conjugated by Hadamards [72]. Therefore we are not the first to observe that discrete gate sets can still have beyond-classical computation power.

[2]In particular because the input state is not a stabilizer state, and the output measurements are not stabilizer measurements.

error in each output probability) by a classical computer, unless the polynomial hierarchy collapses. Furthermore, we provide a *complete classification* of the power of CCCs according the choice of U. In particular, we show that *any* U which is non-Clifford[3] suffices to perform hard (exact) sampling problems with CCCs. This result can be seen as progress towards classifying the computational complexity of restricted gate sets (i.e. Conjecture 2.2.1), which is a challenging open problem [63, 66, 214, 17].

Second, we show that under an additional complexity-theoretic conjecture, that classical computers cannot efficiently simulate U-CCCs to constant error in total variation distance. This is a more realistic model of error for noisy error-corrected quantum computations. Of course even achieving constant total variation distance error as the number of qubits increases will require error-correction. However the particular parameters of our hardness result - namely tolerance to total variation distance 1/64 (vs 1/192 for IQP)- may make our results more achievable without error correction for small circuits. Note that the conjecture assumed to achieve these results is an average-case hardness conjecture, as in [11, 74, 114, 194]. It essentially states that for most Clifford circuits V and most one-qubit unitaries U, most of the output probabilities of the CCC $U^{\otimes n}V(U^{\dagger})^{\otimes n}$ are #P-hard to compute to multiplicative error. We prove that this conjecture is true in the worst case - in fact for all non-Clifford U, there exists a V such that some outputs are #P-hard to compute to multiplicative error. However proving this hardness extends to the average case remains open. Unfortunately recent results of Aaronson and Chen [15] imply that proving this conjecture would require non-relativizing techniques, so we expect this to be a difficult open problem. Note, however, that to the best of our knowledge our conjecture is independent of the conjectures used to establish other quantum advantage results such as boson sampling [11] or IQP [74, 73]. Therefore it can be seen as establishing an alternative basis for a belief in quantum advantage over classical computation.

One motivation for studying CCCs is that they might admit a simpler fault-tolerant implementation than universal quantum computing. It is well-known that many stabilizer error-correcting codes, such as the 5-qubit and 7-qubit codes [178, 104, 224], admit transversal Clifford operations [132]. That is, performing fault-tolerant Clifford operations on the encoded logical qubits can be done in a very simple manner – by simply performing the corresponding Clifford operation on the physical qubits. This is manifestly fault-tolerant, in that an error on one physical qubit does not "spread" to more than 1 qubit when applying the gate. In contrast, performing non-Clifford operations fault-tolerantly on such codes requires substantially larger (and non-transversal) circuits - and therefore the non-tranversal operations are often the most difficult to implement fault-tolerantly. The challenge in fault-tolerantly implementing CCCs therefore lies in performing the initial state preparation and measurement. Initial preparation of non-stabilizer states in these codes is equivalent to the challenge of producing magic states, which are already known to boost Clifford circuits to universality using adaptive Clifford Circuits [70, 69] (in contrast our construction would only need non-adaptive Clifford circuits with magic states). Likewise, measuring in a non-Clifford basis would require performing non-Clifford one-qubit gates prior to fault-tolerant measurement in the computational basis. Therefore the state preparation/measurement would be the challenging part of fault-tolerantly implementing CCCs in codes with transversal Cliffords. It remains open if there exists a code with tranversal conjugated Cliffords[4] and easy

[3]Note that a Z rotation at the beginning or end of a CCC does not affect output probabilities, so the more precise statement is made up to appending a Z rotation to U. See Theorem 5.3.1 for details.

[4]Of course one can always "rotate" a code with transversal Clifford operations to obtain a code with transversal conjugated Cliffords. If the code previously had logical states $|0\rangle_L, |1\rangle_L$, then by setting the states

preparation and measurement in the required basis. Such a code would not be ruled out by the Eastin-Knill Theorem [106] (which states that the set of transversal gates must be discrete for all codes which correct arbitrary one qubit errors).

5.1.2 Proof Techniques

To prove these results, we use several different techniques.

Proof Techniques: classification of exact sampling hardness

To prove exact (or multiplicative) sampling hardness for CCCs for essentially all non-Clifford U, we use the techniques of Section 2.2.3. Our proof works by showing that postselecting such circuits - allows them to perform universal quantum computation. Hardness then follows from known techniques [6, 72, 11].

One technical subtlety that we face in this proof, which is not present in other results, is that our postselected gadgets perform operations which are not closed under inversion. As discussed in Section 2.2.3, this means one cannot use the Solovay-Kitaev theorem to change quantum gate sets [99]. This is a necessary step in the proof that PostBQP = PP [6], which is a key part of the hardness proof (see Section 2.2.3). Fortunately, it turns out that we can get away without inverses due to a recent "Inverse-free Solovay-Kitaev theorem" of Sardharwalla *et al.* [212], which discards the needs for inverses if the gate set contains the Paulis. Our result would have been much more difficult to obtain without this prior result.

A further difficulty in the classification proof is that the postselection gadgets we derive to not work for all non-Clifford U. In general, most postselection gadgets give rise to non-unitary operations, and for technical reasons we need to work with unitary postselection gadgets to apply the results of [212]. Therefore, instead we use several different gadgets which cover different portions of the parameter space of U's. Our initial proof of this fact used a total of seven postselection gadgets found by hand. We later simplified this to two postselection gadgets by writing a computer program to brute-force search for nice gadgets using Christopher Granade and Ben Criger's QuaEC package. We include this simplified proof in this chapter.

Proof techniques: additive error

To prove hardness of simulation to additive error, we follow the techniques of [11, 74, 114, 194]. In these works, to show hardness of sampling from some probability distribution with additive error, one combines three different ingredients. The first is anti-concentration - showing that for these circuits, the output probabilities in some large set T are somewhat large. Second, one uses Markov's inequality to argue that, since the simulation error sums to ε, on some other large set of output probabilities S, the error must be below a constant multiple of the average. If S and T are both large, they must have some intersection - and on this intersection $S \cap T$, the imagined classical simulation is not only a simulation to additive error, but also to multiplicative error as well (since the output probability in question is above some minimum). Therefore a simulation to some amount ε of additive error implies a multiplicative simulation to the output probabilities on a constant fraction of the outputs.

$|0\rangle'_L = U_L^\dagger |0\rangle_L$ and $|1\rangle'_L = U_L^\dagger |1\rangle_L$, one obtains a code in which the conjugated Clifford gates (conjugated by U) are transversal. However having the ability to efficiently fault-tolerantly prepare $|0\rangle_L$ in the old code does not imply the same ability to prepare $|0\rangle'_L$ in the new code.

The impossibility of such a simulation is then obtained by assuming that computing these output probabilities is multiplicatively hard on average. In particular, one assumes the set of outputs T' for which it is #P-hard to compute the output probability is large. Therefore one has a multiplicative approximation of a #P-hard quantity in the set $S \cap T \cap T'$ - which leads to a collapse of the polynomial hierarchy by known techniques [11, 72].

We follow this technique to show hardness of sampling with additive error. In our case, the anticoncentration theorem follows from the fact that the Clifford group is a "3-design" [241, 247] - i.e. a random Clifford circuit behaves equivalently to a random unitary up to its third moment - and therefore must anticoncentrate, as a random unitary does (the fact that unitary designs anticoncentrate was also shown independently by several groups [145, 185, 151]). This is similar to the hardness results for IQP [74] and DQC_1 [194], in which the authors also prove their corresponding anticoncentration theorems. In contrast it is open to prove the anticoncentration theorem used for Boson Sampling [11]. Therefore the only assumption needed is the hardness-on-average assumption.

We also show that our hardness of average assumption is true in for worst case inputs. This result follows from combining known facts about BQP with the classification theorem for exact sampling hardness.

5.1.3 Relation to other works on modified Clifford circuits

While we previously discussed the relation of our results to prior work on sampling problems, here we compare our results to prior work on Clifford circuits. We are not the first to consider the power of modified Clifford circuits. Jozsa and van den Nest [163] and Koh [173], categorized the computational power of a number of modified versions of Clifford circuits. The closest related result is the statement in [163] that if the input state to a Clifford circuit is allowed to be an arbitrary tensor product of one-qubit states, then such circuits cannot be efficiently classically simulated unless the polynomial hierarchy collapses. Their hardness result uses states of the form $|0\rangle^{\otimes n/2}|\alpha\rangle^{\otimes n/2}$, where $|\alpha\rangle = \cos(\pi/8)|0\rangle + i\sin(\pi/8)|1\rangle$ is a magic state. They achieve postselected hardness via using magic states to perform T gates, using a well-known construction (see e.g. [70]). So in their construction there are different input states on different qubits. In contrast, our result requires the same input state on every qubit - as well as measurement in that basis at the end of the circuit. This ensures our modified circuit can be interpreted as the action of a discrete gate set.

5.2 Preliminaries

We denote the single-qubit *Pauli matrices* by $X = \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $Y = \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $Z = \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The ± 1-eigenstates of Z are denoted by $|0\rangle$ and $|1\rangle$ respectively.

The *rotation operator* about an axis $t \in \{x, y, z\}$ with an angle $\theta \in \mathbb{R}$ is

$$R_t(\theta) = e^{-i\theta\sigma_t/2} = \cos(\theta/2)I - i\sin(\theta/2)\sigma_t. \tag{5.1}$$

We will use the fact that any single-qubit unitary operator U can be written as

$$U = e^{i\alpha}R_z(\phi)R_x(\theta)R_z(\lambda), \tag{5.2}$$

where $\alpha, \phi, \theta, \lambda \in \mathbb{R}$ [199].

For linear operators A and B, we write $A \propto B$ to mean that there exists $\alpha \in \mathbb{C}\backslash\{0\}$ such that $A = \alpha B$. For linear operators, vectors or complex numbers a and b, we write $a \sim b$ to mean that a and b differ only by a global phase, i.e. there exists $\theta \in \mathbb{R}$ such that $a = e^{i\theta} b$. For any subset $S \subseteq \mathbb{R}$ and real number $k \in \mathbb{R}$, we write kS to refer to the set $\{kn : n \in S\}$. For example, $k\mathbb{Z} = \{kn : n \in \mathbb{Z}\}$. We denote the set of odd integers by $\mathbb{Z}_{odd}$. We denote the complement of a set S by S^c.

5.2.1 Clifford circuits and conjugated Clifford circuits

The n-qubit *Pauli group* $\mathcal{P}_n$ is the set of all operators of the form $i^k P_1 \otimes \ldots \otimes P_n$, where $k \in \{0, 1, 2, 3\}$ and each P_j is a Pauli matrix. The n-qubit *Clifford group* is the normalizer of $\mathcal{P}_n$ in the n-qubit unitary group $\mathcal{U}_n$, i.e. $\mathcal{C}_n = \{U \in \mathcal{U}_n : U\mathcal{P}_n U^\dagger = \mathcal{P}_n\}$.

The elements of the Clifford group, called *Clifford operations*, have an alternative characterization: an operation is a Clifford operation if and only if it can be written as a circuit comprising the following gates, called *basic Clifford gates*: *Hadamard*, *$\pi/4$ phase*, and *controlled-Z* gates, whose matrix representations in the computational basis are $H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, $S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$, and $CZ = \mathrm{diag}(1, 1, 1, -1)$ respectively. An example of a non-Clifford gate is the T gate, whose matrix representation is $T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$. We denote the group generated by the single-qubit Clifford gates by $\langle S, H \rangle$.

We will make use of the following fact about Clifford operations.

Fact 1. *$R_z(\phi)$ is a Clifford operation if and only if $\phi \in \frac{\pi}{2}\mathbb{Z}$.*

A *Clifford circuit* is a circuit that consists of computational basis states being acted on by the basic Clifford gates, before being measured in the computational basis. Without loss of generality, we may assume that the input to the Clifford circuit is the all-zero state $|0\rangle^{\otimes n}$. We define conjugated Clifford circuits (CCCs) similarly to Clifford circuits, except that each basic Clifford gate G is replaced by a conjugated basic Clifford gate $(U^{\otimes k})^\dagger g U^{\otimes k}$, where $k = 1$ when $g = H, S$ and $k = 2$ when $g = CZ$. In other words,

Definition 5.2.1. Let U be a single-qubit unitary gate. A U-conjugated Clifford circuit (U-CCC) on n qubits is defined to be a quantum circuit with the following structure:

1. Start with $|0\rangle^{\otimes n}$.
2. Apply gates from the set $\{U^\dagger H U, U^\dagger S U, (U^\dagger \otimes U^\dagger) CZ (U \otimes U)\}$.
3. Measure each qubit in the computational basis.

Because the intermediate U and $U^\dagger$ gates cancel, we may equivalently describe a U-CCC as follows:

1. Start with $|0\rangle^{\otimes n}$.
2. Apply $U^{\otimes n}$.
3. Apply gates from the set $\{H, S, CZ\}$.
4. Apply $(U^\dagger)^{\otimes n}$.
5. Measure each qubit in the computational basis.

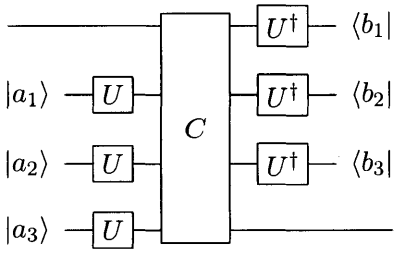
5.2.2 Postselection gadgets

Our results involve the use of postselection gadgets to simulate unitary operations. In this section, we introduce some terminology to describe these gadgets.

Definition 5.2.2. Let U be a single-qubit operation. Let $k, l \in \mathbb{Z}^+$ with $k > l$. A *k-to-l U-CCC postselection gadget* G is a postselected circuit fragment that performs the following procedure on an l-qubit system:

1. Introduce a set T of $(k-l)$ ancilla registers in the state $|a_1 \dots a_{k-l}\rangle$, where $a_1 \dots a_{k-l} \in \{0,1\}^{k-l}$.
2. Apply $U^{\otimes(k-l)}$ to the set T of registers.
3. Apply a k-qubit Clifford operation C to both the system and ancilla.
4. Choose a subset S of $(k-l)$ registers and apply $(U^\dagger)^{\otimes(k-l)}$ to S.
5. Postselect on the subset S of qubits being in the state $|b_1 \dots b_{k-l}\rangle$, where $b_1 \dots b_{k-l} \in \{0,1\}^{k-l}$.

An example of a 4-to-1 U-CCC postselection gadget is the circuit fragment described by the following diagram:



Let G be a U-CCC postselection gadget as described in Definition 5.2.2. The *action* $A(G)$ (also denoted A_G) of G is defined to be the linear operation that it performs, i.e.

$$A(G) = A_G = \langle b_1 \dots b_l|_S \left(\prod_{i \in S} U_i^\dagger\right) C \left(\prod_{i \in T} U_i\right) |a_1 \dots a_l\rangle_T, \tag{5.3}$$

and the *normalized action* of G, when it exists, is

$$\tilde{A}_G = \frac{A_G}{(\det A_G)^{2^{-l}}}. \tag{5.4}$$

Note that the above normalization is chosen so that $\det \tilde{A}_G = 1$.

We say that a U-CCC postselection gadget G is *unitary* if there exists $\alpha \in \mathbb{C}\backslash\{0\}$ and a unitary operator U such that $A_G = \alpha U$. It is straightforward to check that the following are equivalent conditions for gadget unitarity.

Lemma 5.2.3. *A U-CCC postselection gadget G is unitary if and only if either one of the following holds:*

1. *There exists $\gamma > 0$ such that $A_G^\dagger A_G = \gamma I$,*

2. $\tilde{A}_G^\dagger \tilde{A}_G = I$, *i.e.* $\tilde{A}_G$ *is unitary.*

Similarly, we say that a U-CCC postselection gadget G is *Clifford* if there exists $\alpha \in \mathbb{C}\backslash\{0\}$ and a Clifford operator U such that $A_G = \alpha U$. The following lemma gives a necessary condition for a gadget to be Clifford.

Lemma 5.2.4. *If G is a Clifford U-CCC postselection gadget, then*

$$A_G X A_G^\dagger \propto X \text{ or } A_G X A_G^\dagger \propto Y \text{ or } A_G X A_G^\dagger \propto Z, \tag{5.5}$$

and

$$A_G Z A_G^\dagger \propto X \text{ or } A_G Z A_G^\dagger \propto Y \text{ or } A_G Z A_G^\dagger \propto Z. \tag{5.6}$$

Proof. If G is a Clifford U-CCC postselection gadget, then there exists $\alpha \in \mathbb{C}\backslash\{0\}$ and a Clifford operation C such that $A_G = \alpha C$. Since C is Clifford, $CXC^\dagger$ is a Pauli operator. But $CXC^\dagger \not\sim I$, otherwise, $X \sim I$, which is a contradiction. Hence, $CXC^\dagger \sim X$ or Y or Z, which implies Eq. (5.5). The proof of Eq. (5.6) is similar, with X replaced with Z. □

5.3 Weak simulation of CCCs with multiplicative error

5.3.1 Classification results

In this section, we classify the hardness of weakly simulating U-CCCs as we vary U. As we shall see, it turns out that the classical simulation complexities of the U-CCCs associated with this notion of simulation are all of the following two types: the U-CCCs are either efficiently simulable, or are hard to simulate to constant multiplicative error unless the polynomial hierarchy collapses. To facilitate exposition, we will introduce the following terminology to describe these two cases: Let $\mathcal{C}$ be a class of quantum circuits. We say that $\mathcal{C}$ is in PWEAK if it is efficiently simulable in the weak sense by a classical computer. We say that $\mathcal{C}$ is PH-*supreme* (or that it exhibits PH-*supremacy*) if it satisfies the property that if $\mathcal{C}$ is efficiently simulable in the weak sense by a classical computer to constant multiplicative error, then the polynomial hierarchy (PH) collapses.

The approach we take to classifying the U-CCCs is to decompose each U into the form given by Eq. (5.2),

$$U = e^{i\alpha} R_z(\phi) R_x(\theta) R_z(\lambda), \tag{5.7}$$

and study how the classical simulation complexity changes as we vary α, ϕ, θ and λ. Two simplifications can immediately be made. First, the outcome probabilities of the U-CCC are independent of α, since α appears only in a global phase. Second, the probabilities are also independent of λ. To see this, note that the outcome probabilities are all of the form:

$$|\langle b|R_z(-\lambda)^{\otimes n} V R_z(\lambda)^{\otimes n}|0\rangle|^2 = |\langle b|V|0\rangle|^2, \tag{5.8}$$

which is independent of λ. In the above expression, $b \in \{0,1\}^n$ and

$$V = R_x(-\theta) R_z(-\phi) C R_z(\phi) R_x(\theta)$$

for some Clifford circuit C. The equality follows from the fact that the computational basis states are eigenstates of $R_z(\lambda)^{\otimes n}$ with unit-magnitude eigenvalues.

ϕ \ θ	$\pi\mathbb{Z}$	$\frac{\pi}{2}\mathbb{Z}_{odd}$	$\left(\frac{\pi}{2}\mathbb{Z}\right)^c$
$\frac{\pi}{2}\mathbb{Z}$	PWEAK (i, ii)	PWEAK (ii)	PH-supreme (iv)
$\left(\frac{\pi}{2}\mathbb{Z}\right)^c$	PWEAK (i)	PH-supreme (iii)	PH-supreme (iv)

Table 5.1: Complete complexity classification of U-CCCs (where $U = R_z(\phi)R_x(\theta)$) with respect to weak simulation, as we vary ϕ and θ. The roman numerals in parentheses indicate the parts of Theorem 5.3.1 that are relevant to the corresponding box. All U-CCCs are either in PWEAK (i.e. can be efficiently simulated in the weak sense) or PH-supreme (i.e. cannot be simulated efficiently in the weak sense, unless the polynomial hierarchy collapses.)

Hence, to complete the classification, it suffices to just restrict our attention to the two-parameter family $\{R_z(\phi)R_x(\theta)\}_{\phi,\theta}$ of unitaries. Our main result is the following theorem (see Table 5.1 for a summary):

Theorem 5.3.1. *Let $U = R_z(\phi)R_x(\theta)$, where $\phi, \theta \in \mathbb{R}$. Then*

- *U-CCCs are in* PWEAK, *if*

  (i) $\phi \in \mathbb{R}$ and $\theta \in \pi\mathbb{Z}$, or

  (ii) $\phi \in \frac{\pi}{2}\mathbb{Z}$ and $\theta \in \frac{\pi}{2}\mathbb{Z}$.

- *U-CCCs are* PH*-supreme, if*

  (iii) $\phi \notin \frac{\pi}{2}\mathbb{Z}$ and $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$, or

  (iv) $\theta \notin \frac{\pi}{2}\mathbb{Z}$.

We defer the proof of Theorem 5.3.1 to Sections 5.3.2 and 5.3.3. For now, we will derive a consequence of the theorem.

Corollary 5.3.2. *Let U be a single-qubit unitary operator. Consider the following two statements:*

(A) U-CCC is in PWEAK.

(B) There exists a single-qubit Clifford operator $C \in \langle S, H \rangle$ and $\lambda \in \mathbb{R}$ such that

$$U \sim CR_z(\lambda). \tag{5.9}$$

Then,

1. *(B) implies (A).*
2. *If the polynomial hierarchy is infinite, then (A) implies (B).*

Proof.

1. Since $R_z(\lambda)|0\rangle \sim |0\rangle$, it follows that for any C, $CR_z(\lambda)$-CCCs have the same outcome probabilities as C-CCCs. But C-CCCs are efficiently simulable, by the Gottesman-Knill Theorem, since $C \in \langle S, H\rangle$. Hence, U-CCCs are in $PWEAK$.

2. Let U be such that U-CCCs are in $PWEAK$. Using the decomposition in Eq. (5.2), write $U = e^{i\alpha}R_z(\phi)R_x(\theta)R_z(\lambda)$. Since we assumed that the polynomial hierarchy is infinite, Theorem 5.3.1 implies that

   (a) $\theta \in \pi\mathbb{Z}$, or

   (b) $\theta \in \frac{\pi}{2}\mathbb{Z}$ and $\phi \in \frac{\pi}{2}\mathbb{Z}$.

   In Case (a), $\theta \in 2\pi\mathbb{Z}$ or $\pi\mathbb{Z}_{odd}$. If $\theta \in 2\pi\mathbb{Z}$, then

   $$U \sim R_z(\phi)R_x(2\pi\mathbb{Z})R_z(\gamma) = I.R_z(\phi+\gamma),$$

   which is of the form given by Eq. (5.9). If $\pi\mathbb{Z}_{odd}$, then

   $$U \sim R_z(\phi)R_x(\pi\mathbb{Z}_{odd})R_z(\gamma) \sim R_z(\phi)XR_z(\gamma) = XR_z(\gamma-\phi),$$

   which is again of the form given by Eq. (5.9).

   In Case (b),

   $$\begin{aligned} U &\in e^{i\alpha}R_z(\pi\mathbb{Z}/2)R_x(\pi\mathbb{Z}/2)R_z(\gamma) \\ &= e^{i\alpha}R_z(\pi\mathbb{Z}/2)HR_z(\pi\mathbb{Z}/2)HR_z(\gamma). \end{aligned} \tag{5.10}$$

   But the elements of $R_z(\pi\mathbb{Z}/2)$ are of the form S^i, for $i \in \mathbb{Z}$, up to a global phase. Therefore, $R_z(\pi\mathbb{Z}/2)HR_z(\pi\mathbb{Z}/2)H$ is Clifford, and U is of the form Eq. (5.9).

□

Hence, Corollary 5.3.2 tells us that under the assumption that the polynomial hierarchy is infinite, U-CCCs can be simulated efficiently (in the weak sense) if and only if $U \sim CR_z(\lambda)$ for some single qubit Clifford operator C, i.e. if U is a Clifford operation times a Z-rotation.

5.3.2 Proofs of efficient classical simulation

In this section, we prove Cases (i) and (ii) of Theorem 5.3.1.

Proof of Case (i): $\phi \in \mathbb{R}$ and $\theta \in \pi\mathbb{Z}$

Theorem 5.3.3. *Let $U = R_z(\phi)R_x(\theta)$. If $\phi \in \mathbb{R}$ and $\theta \in \pi\mathbb{Z}$, then U-CCCs are in* PWEAK.

Proof. First, we consider the case where $\theta \in 2\pi\mathbb{Z}$. In this case, $U = R_z(\phi)$, and the amplitudes of the U-CCC can be written as

$$\langle y|R_z(-\phi)^{\otimes n}CR_z(\phi)^{\otimes n}|x\rangle \sim \langle y|C|x\rangle \tag{5.11}$$

for some Clifford operation C and computational basis states $|x\rangle$ and $|y\rangle$. By the Gottesman-Knill Theorem, these U-CCCs can be efficiently weakly simulated.

Next, we consider the case where $\theta \in \pi\mathbb{Z}_{odd}$. In this case, $U = R_z(\phi)R_x(\pi) \sim R_z(\phi)X$, and the amplitudes of the U-CCC can be written as

$$\langle y|XR_z(-\phi)^{\otimes n}CR_z(\phi)^{\otimes n}X|x\rangle \sim \langle \bar{y}|C|\bar{x}\rangle \tag{5.12}$$

for some Clifford operation C and computational basis states $|x\rangle$ and $|y\rangle$, where $\bar{z} = 1 - z$. By the Gottesman-Knill Theorem, these U-CCCs can be efficiently weakly simulated.

Putting the above results together, we get that U-CCCs are in PWEAK. □

Proof of Case (ii): $\phi \in \frac{\pi}{2}\mathbb{Z}$ and $\theta \in \frac{\pi}{2}\mathbb{Z}$

Theorem 5.3.4. *Let $U = R_z(\phi)R_x(\theta)$. If $\phi \in \frac{\pi}{2}\mathbb{Z}$ and $\theta \in \frac{\pi}{2}\mathbb{Z}$, then U-CCCs are in* PWEAK.

Proof. The elements of $R_z(\frac{\pi}{2}\mathbb{Z})$ are of the form S^i, where $i \in \mathbb{Z}$, up to a global phase. Therefore, $U = R_z(\phi)R_x(\theta) = R_z(\phi)HR_z(\theta)H$ is a Clifford operation, and so, the U-CCCs consist of only Clifford gates. By the Gottesman-Knill Theorem, these U-CCCs can be be efficiently (weakly) simulated. □

5.3.3 Proofs of hardness

In this section, we prove Cases (iii) and (iv) of Theorem 5.3.1. We start by proving a lemma that will be useful for the proofs of hardness.

Lemma 5.3.5. *(Sufficient condition for* PH*-supremacy) Let U be a single-qubit gate. If there exists a unitary non-Clifford U-CCC postselection gadget G, then U-CCCs are* PH*-supreme.*

Proof. Suppose such a gadget G exists. Then, by the inverse-free Solovay-Kitaev Theorem of Sardharwalla *et al.* [212], using polynomially many gates from the set G, H, S one can compile any desired one-qubit unitary V to inverse exponential accuracy (since in particular $\langle H, S\rangle$ contains the Paulis). In particular, since any three-qubit unitary can be expressed as a product of a constant number of CNOTs and one-qubit unitaries, one can compile any gate in the set {CCZ, Controlled-H, all one-qubit gates } to inverse exponential accuracy with polynomial overheard.

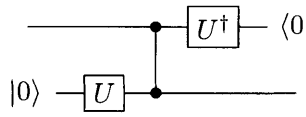
In his proof that PostBQP = PP, Aaronson showed that postselected poly-sized circuits of the above gates can compute any language in PP [6]. Furthermore, as his postselection succeeds with inverse exponential probability, compiling these gates to inverse exponential accuracy is sufficient for performing arbitrary PP computations.

Hence, by using polynomially many gadgets for G , CNOT, H and S, one can compile Aaronson's circuits[5] for computing PP to inverse exponential accuracy, and hence these circuits can compute PP-hard problems. PH-supremacy then follows from the techniques of [72, 11]. Namely, a weak simulation of such circuits with constant multiplicative error would place $\mathsf{PP} \subseteq \mathsf{BPP}^{\mathsf{NP}} \subseteq \Delta_3$ by Stockmeyer counting, and hence by Toda's theorem this would result in the collapse of PH. For details of the argument we refer the reader to Section 2.2.3. □

[5]More specifically, we compile the circuit given by $(U^\dagger)^{\otimes n}$, then Aaronson's circuit, then $U^{\otimes n}$, as we need to cancel the U's at the beginning and the $U^\dagger$s at the end in order to perform Aaronson's circuit which starts and measures in the computational basis. However as the $U, U^\dagger$ are one-qubit gates, one can cancel them to inverse exponential accuracy using our gates, and hence this construction suffices.

Proof of Case (iii): $\phi \notin \frac{\pi}{2}\mathbb{Z}$ **and** $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$

Let $U = R_z(\phi)R_x(\theta)$. Consider the following U-CCC postselection gadget:

$I(\phi,\theta) =$ 

(5.13)

We now prove some properties about $I(\phi,\theta)$.

Theorem 5.3.6.

1. *The action of* $I(\phi,\theta)$ *is*

$$A_{I(\phi,\theta)} = \begin{pmatrix} \cos^2\frac{\theta}{2} & \frac{i}{2}\sin\theta\, e^{-i\phi} \\ -\frac{i}{2}\sin\theta\, e^{i\phi} & -\sin^2\frac{\theta}{2} \end{pmatrix}. \tag{5.14}$$

2. $I(\phi,\theta)$ *is a unitary gadget if and only if* $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$. *When* $I(\phi,\theta)$ *is unitary,*

$$\tilde{A}_{I(\phi,\theta)} = \frac{i}{\sqrt{2}}\begin{pmatrix} 1 & i(-1)^k e^{-i\phi} \\ -i(-1)^k e^{i\phi} & -1 \end{pmatrix}, \tag{5.15}$$

   where $k = \frac{\theta}{\pi} - \frac{1}{2}$.

3. $I(\phi,\theta)$ *is a Clifford gadget if and only if* $\phi \in \frac{\pi}{2}\mathbb{Z}$ *and* $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$.

4. $I(\phi,\theta)$ *is a unitary non-Clifford gadget if and only if* $\phi \notin \frac{\pi}{2}\mathbb{Z}$ *and* $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$.

Proof.

1. By direct calculation.

2. By Eq. (5.14),

$$A^\dagger_{I(\phi,\theta)}A_{I(\phi,\theta)} = \begin{pmatrix} \cos^2\frac{\theta}{2} & \frac{i}{4}\sin(2\theta)e^{-i\phi} \\ -\frac{i}{4}\sin(2\theta)e^{i\phi} & \sin^2\frac{\theta}{2} \end{pmatrix}. \tag{5.16}$$

   If $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$, then $A^\dagger_{I(\phi,\theta)}A_{I(\phi,\theta)} = \frac{1}{2}I$, which implies that $I(\phi,\theta)$ is a unitary gadget, by Lemma 5.2.3. Conversely, assume that $I(\phi,\theta)$ is a unitary gadget. Suppose that $\theta \notin \frac{\pi}{2}\mathbb{Z}_{odd}$. Then $\sin(2\theta) \neq 0$, which implies that $A^\dagger_{I(\phi,\theta)}A_{I(\phi,\theta)} \not\propto I$, which is a contradiction. Hence, $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$.

   Next, $k = \frac{\theta}{\pi} - \frac{1}{2}$ implies that $\theta = \frac{\pi}{2}(2k+1)$. Since $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$, it follows that $k \in \mathbb{Z}$. Then $\sin\theta = (-1)^k$, $\cos^2\frac{\theta}{2} = \frac{1}{2}$ and $\sin^2\frac{\theta}{2} = \frac{1}{2}$. Hence,

$$A_{I(\phi,\theta)} = \begin{pmatrix} \frac{1}{2} & \frac{i}{2}(-1)^k e^{-i\phi} \\ -\frac{i}{2}(-1)^k e^{i\phi} & -\frac{1}{2} \end{pmatrix}. \tag{5.17}$$

   Hence, $\det A_{I(\phi,\theta)} = -\frac{1}{2}$. Plugging this and Eq. (5.17) into Eq. (5.4) gives Eq. (5.15).

3. ($\Leftarrow$) Let $\phi \in \frac{\pi}{2}\mathbb{Z}$ and $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$. Write $\phi = \frac{\pi}{2}l$ and $\theta = \frac{\pi}{2}(2k+1)$. Then, by Eq. (5.15),

$$\tilde{A}_{I(\phi,\theta)} = \frac{i}{\sqrt{2}}\begin{pmatrix} 1 & i^{1+2k+3l} \\ i^{3+2k+l} & -1 \end{pmatrix}. \tag{5.18}$$

Now, it is straightforward to check that for all $k, l \in \mathbb{Z}$, $\tilde{A}_{I(\phi,\theta)} X \tilde{A}^{\dagger}_{I(\phi,\theta)} \in \{-X, Z, -Z\}$ and $\tilde{A}_{I(\phi,\theta)} Z \tilde{A}^{\dagger}_{I(\phi,\theta)} \in \{-Y, X, Y, -X\}$. This shows that $\tilde{A}_{I(\phi,\theta)}$ maps the Pauli group to itself, under conjugation, which implies that $\tilde{A}_{I(\phi,\theta)}$ is Clifford.

($\Rightarrow$) Assume that $I(\phi,\theta)$ is a Clifford gadget. Suppose that $\phi \notin \frac{\pi}{2}\mathbb{Z}$ or $\theta \notin \frac{\pi}{2}\mathbb{Z}_{odd}$. But $I(\phi,\theta)$ is unitary, and hence, $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$. So $\phi \notin \frac{\pi}{2}\mathbb{Z}$. By Lemma 5.2.4, $\tilde{A}_{I(\phi,\theta)} X \tilde{A}^{\dagger}_{I(\phi,\theta)} \sim X$ or Y or Z. But, as we compute,

$$\tilde{A}_{I(\phi,\theta)} X \tilde{A}^{\dagger}_{I(\phi,\theta)} = \begin{pmatrix} (-1)^k \sin\phi & -e^{-i\phi}\cos\phi \\ -e^{i\phi}\cos\phi & -(-1)^k \sin\phi \end{pmatrix}. \tag{5.19}$$

If $\tilde{A}_{I(\phi,\theta)} X \tilde{A}^{\dagger}_{I(\phi,\theta)} \sim X$ or Y, then $\sin\phi = 0$, which is a contradiction, since $\phi \notin \frac{\pi}{2}\mathbb{Z}$. Hence, $\tilde{A}_{I(\phi,\theta)} X \tilde{A}^{\dagger}_{I(\phi,\theta)} \sim Z$, which implies that $\cos\phi = 0$. But this also contradicts $\phi \notin \frac{\pi}{2}\mathbb{Z}$. Hence, $\phi \in \frac{\pi}{2}\mathbb{Z}$ and $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$.

4. Follows from Parts 2 and 3 of Theorem 5.3.6.

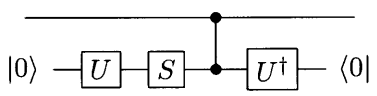
□

Theorem 5.3.7. *Let $U = R_z(\phi)R_x(\theta)$. If $\phi \notin \frac{\pi}{2}\mathbb{Z}$ and $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$, then U-CCCs are* PH*-supreme.*

Proof. By Theorem 5.3.6, when $\phi \notin \frac{\pi}{2}\mathbb{Z}$ and $\theta \in \frac{\pi}{2}\mathbb{Z}_{odd}$, then $I(\phi,\theta)$ is a unitary non-Clifford U-CCC postselection gadget. Hence, by Lemma 5.3.5, U-CCCs are PH-supreme. □

Proof of Case (iv): $\theta \notin \frac{\pi}{2}\mathbb{Z}$

Let $U = R_z(\phi)R_x(\theta)$. Consider the following U-CCC postselection gadget:

$J(\phi,\theta) =$ 

(5.20)

We now prove some properties about $J(\phi,\theta)$.

Theorem 5.3.8.

1. *The action of $J(\phi,\theta)$ is*

$$\begin{aligned} A_{J(\phi,\theta)} &= \frac{1}{\sqrt{2}} e^{-i\frac{\pi}{4}} \begin{pmatrix} i+\cos\theta & 0 \\ 0 & 1+i\cos\theta \end{pmatrix} \\ &= \frac{i}{\sqrt{2}} e^{-i\frac{\pi}{4}} \sqrt{1+\cos^2\theta}\; S^{\dagger} R_z(2\tan^{-1}(\cos\theta)). \end{aligned} \tag{5.21}$$

2. *$J(\phi,\theta)$ is a unitary gadget for all $\theta, \phi \in \mathbb{R}$. The normalized action is*

$$\tilde{A}_{J(\phi,\theta)} \sim S^{\dagger} R_z(2\tan^{-1}(\cos\theta)). \tag{5.22}$$

3. *$J(\phi,\theta)$ is a Clifford gadget if and only if $\theta \in \frac{\pi}{2}\mathbb{Z}$.*

4. *$J(\phi,\theta)$ is a unitary non-Clifford gadget if and only if $\theta \notin \frac{\pi}{2}\mathbb{Z}$.*

Proof.

1. By direct calculation.

2. The determinant of $A_{J(\phi,\theta)}$ is

$$\det A_{J(\phi,\theta)} = \tfrac{1}{2}(1+\cos^2\theta) \neq 0 \tag{5.23}$$

for all θ and ϕ. Hence, $J(\phi,\theta) \propto R_z(2\tan^{-1}(\cos\theta))$ for all θ and ϕ, which implies that $J(\phi,\theta)$ is a unitary gadget for all θ and ϕ.

Hence,

$$\tilde{A}_{J(\phi,\theta)} = \frac{A_{J(\phi,\theta)}}{\sqrt{\det A_{J(\phi,\theta)}}} = ie^{-i\frac{\pi}{4}} S^\dagger R_z(2\tan^{-1}(\cos\theta)).$$

3.

$$\begin{aligned} J(\phi,\theta) \text{ is a Clifford gadget} &\Leftrightarrow S^\dagger R_z(2\tan^{-1}(\cos\theta)) \text{ is Clifford} \\ &\Leftrightarrow R_z(2\tan^{-1}(\cos\theta)) \text{ is Clifford} \\ &\Leftrightarrow 2\tan^{-1}(\cos\theta) \in \tfrac{\pi}{2}\mathbb{Z} \quad \text{by Fact 1} \\ &\Leftrightarrow \cos\theta \in \{0,1,-1\} \\ &\Leftrightarrow \theta \in \tfrac{\pi}{2}\mathbb{Z}. \end{aligned} \tag{5.24}$$

4. Follows from Parts 2 and 3 of Theorem 5.3.8.

□

Theorem 5.3.9. *Let $U = R_z(\phi)R_x(\theta)$. If $\theta \notin \frac{\pi}{2}\mathbb{Z}$, then U-CCCs are* PH*-supreme.*

Proof. By Theorem 5.3.8, when $\theta \notin \frac{\pi}{2}\mathbb{Z}$, then $I(\phi,\theta)$ is a unitary non-Clifford U-CCC postselection gadget. Hence, by Lemma 5.3.5, U-CCCs are PH-supreme. □

5.4 Weak simulation of CCCs with additive error

Here we show how to achieve additive hardness of simulating conjugated Clifford circuits, under additional hardness assumptions. Specifically, we will show that under these assumptions, there is no classical randomized algorithm which given a one-qubit unitary U and a Clifford circuit V, samples the output distribution of V conjugated by U's up to constant ℓ_1 error.

Theorem 5.4.1. *Assuming that PH is infinite and Conjecture 5.4.1, then there is no classical algorithm which given U, V outputs a sampling from the Conjugated Clifford Circuit (U,V) up to additive error $1/64$.*

Conjecture 5.4.1. *For any U which is not equal to a Z-rotation times a Clifford, it is* #P*-hard to approximate $31/32 + o(1)$ fraction of the $p_{y,U,V}$ over the choice of y, V to within multiplicative error $1/2 + o(1)$.*

In order to prove this we'll actually prove a more general theorem described below; the result will then follow from simply setting $a = c = 1/4$, $\varepsilon = 1/64$. One can in general plug in any values they like subject to the constraints; for instance one can weaken the hardness

assumption to a smaller fraction of the $p_{y,U,V}$, or weaken the multiplicative error constant required, if one demands a smaller error in the simulation[6]. These parameters are similar to those appearing in other hardness conjectures; for instance the hardness conjecture for IQP in [74] requires 23/24 of the output probabilities[7] to be hard to compute to error 1/4, in order to get hardness to total variation distance 1/191.

Theorem 5.4.2. *Pick constants $0 < \varepsilon, a, c < 1$ such that $(1-a)^2/2 - c > 0$ and $\frac{2\varepsilon}{ac} < 1$. Then assuming Conjecture 5.4.2, one cannot weakly simulate the distribution D with a randomized classical algorithm with total variation distance error ε, unless the polynomial hierarchy collapses to the third level.*

Conjecture 5.4.2. *For any U which is not equal to a Z-rotation times a Clifford, it is #P-hard to multiplicatively approximate $\left(1 - (1-a)^2/2 + c\right) + o(1)$ fraction of the $p_{y,U,V}$ over the choice of (y, V), up to multiplicative error $\frac{2\varepsilon}{ac} + o(1)$.*

Proof of Theorem 5.4.2. Let V be a Clifford circuit on n qubits, U be a one-qubit unitary which is not a Z-rotation times a Clifford, and $y \in \{0,1\}^n$ be an n-bit string. Define

$$p_{y,U,V} = \left|\langle y|U^{\otimes n} V (U^\dagger)^{\otimes n}|0^n\rangle\right|^2.$$

In other words $p_{y,U,V}$ is the probability of outputting the string y when applying the circuit V conjugated by U's to the all 0's state, and then measuring in the computational basis. Let the corresponding probability distribution on y's given U and V be denoted $D(U,V)$.

Suppose by way of contradiction that there exists a classical poly-time randomized algorithm which given inputs U, V outputs samples from a distribution $D'(U,V)$ such that $\frac{1}{2}|D(U,V) - D'(U,V)|_1 < \varepsilon$. In particular let $q_{y,U,V}$ the probability that $D'(U,V)$ outputs y - i.e. the probability that the simulation outputs y under inputs U, V.

By our simulation assumption, for all U, V we have that $\sum_y |q_{y,U,V} - p_{y,U,V}| \leq 2\varepsilon$. Therefore by Markov's inequality, given our constant $0 < c < 1$, we have that for all U and V there exists a set $S' \subseteq \{0,1\}^n$ of output strings y of size $|S'|/2^n > 1 - c$, such that for all $y \in S'$,

$$|q_{y,U,V} - p_{y,U,V}| \leq \frac{2\varepsilon}{c2^n}.$$

In particular, by averaging over V's, we see that for any U as above, there exists a set $S \subset \{0,1\}^n \times \mathcal{C}$ of pairs (y,V) such that for all $(y,V) \in S$, $|q_{y,U,V} - p_{y,U,V}| \leq \frac{2\varepsilon}{c2^n}$. Furthermore S has measure at least $(1-c)$ over a uniformly random choice of (y,V).

We now show the following anticoncentration lemma (similar theorems were shown independently in [145, 185, 151]):

Lemma 5.4.3. *For any fixed U and y as above, and for any constant $0 < a < 1$, we have that at least $\frac{(1-a)^2}{2}$ fraction of the Clifford circuits V have the property that*

[6] E.g. the setting $a = 1/4$, $c = 1/8$, $\varepsilon = 1/128$ yields hardness to additive error $1/128$ assuming $27/32$ of the $p_{y,U,V}$ are #P-hard to compute to multiplicative error $1/2$

[7] Their conjecture is phrased to say that computing $1/24 + o(1)$ of the output probabilities is a hard task, but this is equivalent to stating that $> 23/24$ of them are #P-hard, as in these proofs one has no control over which elements lie in $S \cap T$.

$$p_{y,U,V} \geq \frac{a}{2^n}$$

We will prove Lemma 5.4.3 shortly. First, we will show why this implies Theorem 5.4.2. In particular, by averaging Lemma 5.4.3 over y's, we see that for any U as above, there exists a set $T \subset \{0,1\}^n \times \mathcal{C}$ of pairs (y,V) such that for all $(y,V) \in T$, $p_{y,U,V} \geq \frac{a}{2^n}$. Furthermore T has measure at least $\frac{(1-a)^2}{2}$ over a uniformly random choice of (y,V). Since we assumed that $(1-a)^2/2 + (1-c) > 1$, then $S \cap T$ must be nonempty, and in particular must contain $(1-a)^2/2 - c$ fraction of the pairs (y,V). On this set $S \cap T$, we have that

$$q_{y,U,V} \leq p_{y,U,V} + \frac{2\varepsilon}{c2^n} = p_{y,U,V} + \frac{2\varepsilon}{ac}\frac{a}{2^n} \leq \left(1 + \frac{2\varepsilon}{ac}\right) p_{y,U,V}$$

and likewise

$$q_{y,U,V} \geq p_{y,U,V} - \frac{2\varepsilon}{c2^n} = p_{y,U,V} - \frac{2\varepsilon}{ac}\frac{a}{2^n} \geq \left(1 - \frac{2\varepsilon}{ac}\right) p_{y,U,V}$$

Since $1 - \frac{2\varepsilon}{ac} > 0$ (which we guaranteed by assumption), $q_{y,U,V}$ is a multiplicative approximation to $p_{y,U,V}$ with multiplicative error $\frac{2\varepsilon}{ac}$ for (y,V) in the set $S \cap T$. The set $S \cap T$ contains at least $(1-a)^2/2 - c$ fraction of the total pairs (y,V).

On the other hand, by Conjecture 5.4.2 we have that $1 - ((1-a)^2/2 - c) + o(1)$ fraction of the $p_{y,U,V}$ are #P-hard to approximate to this level of multiplicative error - call this set T'. Now note that these sets must intersect with measure at least $o(1)$ (say 1/poly). In particular, there exists some (y^*, V^*) in $S \cap T \cap T'$. For this, approximating p_{y,U^*,V^*} to this level of multiplicative error is both #P-hard, and achievable by our simulation algorithm. This collapses PH to the third level by known arguments [11, 72]. In particular, by applying Stockmeyer's approximate counting algorithm [225] to p_{y^*,U,V^*}, one can multiplicatively approximate q_{y^*,U,V^*} to multiplicative error $\frac{1}{\mathsf{poly}}$ in $\mathsf{FBPP}^{\mathsf{NP}}$. But since q_{y^*,U,V^*} is a $\frac{2\varepsilon}{ac}$-approx to p_{y^*,U,V^*}, this is a $\frac{2\varepsilon}{ac} + o(1)$ multiplicative approximation to p_{y^*,U,V^*}. Hence a #P-hard quantity is in $\mathsf{FBPP}^{\mathsf{NP}}$. This collapses PH to the third level by Toda's theorem [232].

To complete our proof of Theorem 5.4.2, we will prove Lemma 5.4.3.

Proof of Lemma 5.4.3. To prove this, we will make use of the fact that the Clifford group is an exact 3-design [241, 247]. The fact that the Clifford group is a 3-design means that for any polynomial p over the variables $\{V_{ij}\}$ and their complex conjugates, which is of degree at most 3 in the V_{ij}'s and degree at most 3 in the V_{ij}^*'s, we have that

$$\frac{1}{|C|}\sum_{V \in C} p(V, V^*) = \int p(V, V^*) \mathrm{d}V$$

where the integral $\mathrm{d}V$ is taken over the Haar measure. In other words, the expectation values of low-degree polynomials in the entries of the matrices are exactly identical to the expectation values over the Haar measure.

In particular, note that $p_{y,U,V}$ is a degree-1 polynomial in the entries of V and their complex conjugates, and $p^2_{y,U,V}$ is a degree-2 polynomial in these variables. Therefore, since the Clifford group is an exact 2-design (which is implied by it being an exact 3-design), we have that for any y and U,

$$\frac{1}{C}\sum_{V\in\mathcal{C}} p_{y,U,V} = \int p_{y,U,V}\mathrm{d}V = \frac{1}{2^n}$$

and

$$\frac{1}{C}\sum_{V\in\mathcal{C}} p^2_{y,U,V} = \int p^2_{y,U,V}\mathrm{d}V = \frac{2}{2^{2n}-1}\left(1-\frac{1}{2^n}\right)$$

where the values of these integrals over the Haar measure are well known - see for instance Appendix D of [149].

Following [74], we now invoke the Paley-Zygmund inequality, which states that:

Fact 2. *Given a parameter $0 < a < 1$, and a non-negative random variable p of finite variance, we have*

$$\Pr[p \geq a\mathbb{E}[p]] \geq (1-a)^2\mathbb{E}[p]^2/\mathbb{E}[p^2]$$

Applying this inequality to the random variable $p_{y,U,V}$ over the choice of the Clifford circuit V, we have that

$$\Pr_V\left[p_{y,U,V} \geq \frac{a}{2^n}\right] \geq (1-a)^2\frac{2^{-2n}}{\frac{2-2^{-n+1}}{2^{2n}-1}} = (1-a)^2\frac{1-2^{-2n}}{2-2^{-n+1}} \geq \frac{(1-a)^2}{2}$$

which implies the claim.

□

This completes the proof of Theorem 5.4.2.

□

5.5 Evidence in favor of hardness conjecture

In Section 5.4, we saw that by assuming an average case hardness conjecture (namely Conjecture 5.4.2), we could show that a weak simulation of CCCs to additive error would collapse the polynomial hierarchy. A natural question is: what evidence do we have that Conjecture 5.4.2 is true?

In this section, we show that the worst-case version of Conjecture 5.4.2 is true. In fact, we show that for any $U \neq CR_Z(\theta)$ for a Clifford C, there exists a Clifford circuit V and an output y such that computing $p_{y,U,V}$ is #P-hard to constant multiplicative error. Therefore certainly *some* output probabilities of CCCs are #P-hard to compute. Conjecture 5.4.2 is merely conjecturing further that *most* of them are hard to compute.

Theorem 5.5.1 (Worst-case version of Conjecture 5.4.2)**.** *For any U which is not equal to a Z-rotation times a Clifford, there exists a Clifford circuit V and string $y \in \{0,1\}^n$ such that it is #P-hard to multiplicatively approximate a $p_{y,U,V}$ to multiplicative error $1/2 - o(1)$.*

Proof. This follows from combining the ideas from the proof of Theorem 5.3.1 with previously known facts about BQP. In particular, we will use the following facts:

1. There exists a uniform family of poly-size BQP[8] circuits C_x where $x \in \{0,1\}^n$ using a gate set with algebraic entries such that computing $|\langle 0^n|C_x|0^n\rangle|^2$ to multiplicative error $1/2$ is #P-hard [74].

2. For any poly-sized quantum circuit C over a gate set with algebraic entries, any non-zero output probability has magnitude at least inverse exponential [176].

3. As shown in the proof of Theorem 5.3.1, for any U which is not a Clifford gate times a Z rotation, there is a postselection gadget G which performs a unitary but non-Clifford one-qubit operation. Furthermore all ancilla qubits in G begin in the state $|0\rangle$.

From these facts, we can now prove the theorem. Let $p = |\langle 0^n|C_x|0^n\rangle|^2$. By Fact 2, the circuit C_x from Fact 1 either has $p = 0$ or $p \geq 2^{-O(n^c)}$ for some constant c. Now suppose we compile the circuit C_x from Fact 1 using Clifford gates plus the postselection gadget G - call this new circuit with postselection C'_x. By Sardharwalla *et al.* [212] we can compile this circuit with accuracy $\varepsilon = 2^{-O(n^c)-100}$ with only polynomial overhead.

Let $\ell \in \{0,1\}^k$ is the string of postselection bits of the circuit C'_x (which without loss of generality are the last bits of the circuit), and let α is the probability that all postselections succeed. Note α is a known and easily calculated quantity, since each postselection gadget is unitary so succeeds with a known constant probability.

Let $p' = |\langle 0^n\ell|C'_x|0^{n+k}\rangle|^2/\alpha$. Then we have that:

- If $p = 0$ then $p' \leq 2^{-O(n^c)-100}$
- If $p \neq 0$ then $p - 2^{-O(n^c)-100} \leq p' \leq p + 2^{-O(n^c)-100}$. Since $p \geq 2^{-O(n^c)}$, this is a multiplicative approximation to p with error 2^{-100}.

Now suppose that one can compute $|\langle 0^n\ell|C'_x|0^{n+k}\rangle|^2$ to multiplicative error γ to be chosen shortly. Then immediately one can compute $p' = |\langle 0^n\ell|C'_x|0^{n+k}\rangle|^2/\alpha$ to the same amount of multiplicative error - call this estimate p''. By the above argument, if $p = 0$ then $p'' < 2^{-O(n^c)-100}(1+\gamma)$. On the other hand if $p > 0$ then $p' > 2^{-O(n^c)}$, so $p'' > 2^{-O(n^c)}(1-\gamma)$. In particular if γ is between $1/100$ and $99/100$ these two cases can be distinguished.

Therefore, if $p'' < 2^{-O(n^c)}$ then we can infer that $p = 0$. If $p'' > 2^{-O(n^c)}(1-\gamma)$, then $p > 0$ so p'' is a γ approximation to p' and hence a $\gamma + 2^{-100} + \gamma 2^{-100}$ approximation to p. In either case we have computed a $\gamma + 2^{-100} + \gamma 2^{-100}$ approximation to p. Therefore, if $\gamma = 1/2 - 2^{-99}$, then we have computed a $1/2$-multiplicative approximation to p, which is #P-hard by Fact 1. Therefore, computing some the probability that the CCC correspoding to C'_x outputs $|0^n\ell\rangle$ to multiplicative error $1/2 - 2^{-99}$ is #P-hard. One can similarly improve this hardness to $1/2 - o(1)$.

□

Given that the worst-case version of Conjecture 5.4.2 is true, a natural question to ask is how difficult it would be to prove the conjecture. To do so would in particular prove quantum advantage over classical computation with realistic error, and merely assuming the polynomial hierarchy is infinite. In some ways this would be stronger evidence for quantum advantage over classical computation than Shor's factoring algorithm, as there are no known negative complexity-theoretic consequences if factoring is contained in P.

Unfortunately, recent work has shown that proving Conjecture 5.4.2 would be a difficult task. Specifically, Aaronson and Chen [15] demonstrated an oracle relative to which PH is

[8] Even IQP suffices here [74].

infinite, but classical computers can efficiently weakly simulate quantum devices to constant additive error. Therefore, any proof which establishes quantum advantage with additive error under the assumption that PH is infinite must be non-relativizing. In particular this implies any proof of Conjecture 5.4.2 would require non-relativizing techniques - in other words it could not remain true if one allows for classical oracle class in the circuit. This same barrier holds for proving the similar average-case hardness conjectures to show advantage for Boson Sampling, IQP, DQC_1, or Fourier sampling. Therefore any proof of Conjecture 5.4.2 would require facts specific to the Clifford group. We leave this as an open problem.

5.6 Measurement-based Quantum Computing Proof of Multiplicative Hardness for CCCs for certain U's

In this section, we will prove a weaker version of Theorem 5.3.1, which we state below, using techniques from Measurement-based quantum computing (MBQC).

Theorem 5.6.1. *CCCs with $U = R_Z(\theta)H$ cannot be efficiently weakly classically simulated to multiplicative error $1/2$ unless the polynomial hierarchy collapses, for any θ which is not an integer multiple of $\pi/8$.*

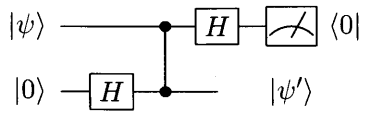
We include it for pedagogical reasons, as it provides a different way of understanding the main theorem using MBQC techniques, and it includes a more detailed walkthrough of the hardness construction. Furthermore, it does not rely on the theorem that the Clifford group plus any non-Clifford element is universal; instead one can directly prove postselected universality by finding a qubit rotation by an irrational multiple of π.

As in the proof of Theorem 5.3.1, we will first show that CCCs can perform universal quantum computation (i.e., the class BQP) under postselection. This first step will make extensive use of ideas from Measurement-Based Quantum Computation [75]. Next, we will show that these circuits can furthermore perform PostBQP under postselection. This extension requires the inverse-free Solovay-Kitaev theorem of Sardharwalla *et al.* [212].

Theorem 5.6.2. *Postselected CCCs can be used to simulate universal quantum computation under the choice of $U = R_Z(\theta)H$, and for any choice of θ other than integer multiples of $\pi/8$.*

Proof. We will first describe the proof without reference to Measurement Based Quantum Computing (MBQC) so as to be understood by the broadest possible audience. We will then summarize the proof in MBQC language for those familiar with the area.

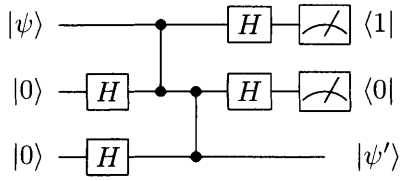
Our proof will make use of four gadgets to show that under postselection, we can perform arbitrary 1-qubit gates in this model. For the first gadget, consider the following quantum circuit:



(5.25)

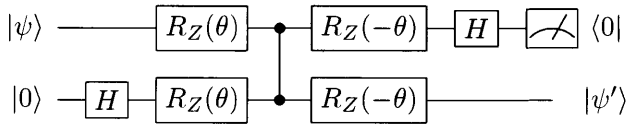
Here the notation $\langle 0|$ denotes that we postselect that measurement outcome on obtaining the state $|0\rangle$, and the two-qubit gate is controlled-Z. This gadget performs teleportation [134]. One can easily calculate that $|\psi'\rangle = H|\psi\rangle$ - in other words, this gadget performs the H gate [134, 72]. Likewise, if one postselects the first outcome to be $|1\rangle$, then the gate performed is

XH. By chaining these gadgets together, one can perform any product of these operations. For instance, the following circuit performs HXH:



The correctness follows from the fact that the order in which quantum measurements are taken is irrelevant. By stringing together n of these, we can perform n gates from the set $\{H, XH\}$. These generate a finite set of one-qubit gates which contain the Paulis.

Now clearly circuits composed of these gadgets do not have the form of Conjugated Clifford circuits with $U = R_Z(\theta)H$. But we can easily correct this by inserting $R_Z(\theta)$'s at the beginning of each line, and $R_Z(-\theta)$'s at the end of each line.

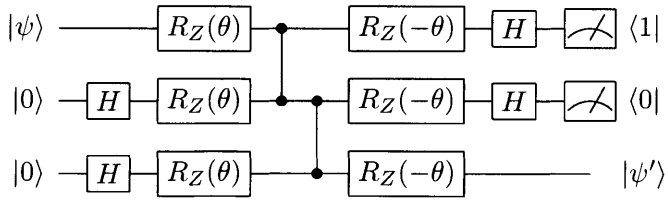


(5.26)

Clearly this is equivalent to our original gadget at the Z rotations commute through and cancel. Now the gadget has the property that

- Every input line begins with $R_Z\theta$, and every output line ends with $R_Z(-\theta)$.
- Every ancillary input begins with $|0\rangle$ then applies $R_Z\theta H$.
- Every ancillary output applies $HR_Z-\theta$ and measures in the computational basis.
- All gates inbetween are Clifford

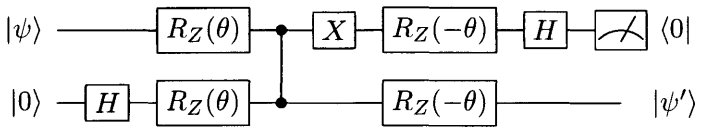
When composing such gadgets, the $R_Z(-\theta)$ at the end of each output line cancels with the $R_Z(\theta)$ at the beginning of each input line. Hence composing gadgets with the above properties will always form a CCC. For instance our prior circuit performing HXH becomes



Thus, by simply replacing our input state $|\psi\rangle$ with the state $H|0\rangle$, and our output state with a Hadamard followed by measurement, this postselected circuit would be simulating the circuit which starts in the state $H|0\rangle$, applies HXH, then applies H and measures. Furthermore, this state will have the form of a CCC. More generally, by stringing n such gadgets together to form a CCC, clearly one can simulate any one-qubit quantum circuit where the initial state is $H|0\rangle$, one performs n gates from the set $\{H, XH\}$, and then applies H and measures.

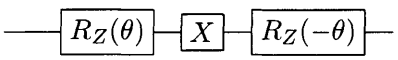
This allows us to simulate one-qubit gates from the set $\{H, XH\}$ with postselected CCC circuits. However, such gates are not universal for a single qubit. In order to show

postselected CCCs can perform universal quantum computation, we will need to find a way to simulate all single qubit gates. To do so, we will consider adding features to our gadget. So far the Clifford part of our CCCs are all commuting; let's consider adding a non-commuting one-qubit gate X to make a new gadget:

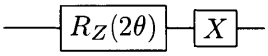


(5.27)

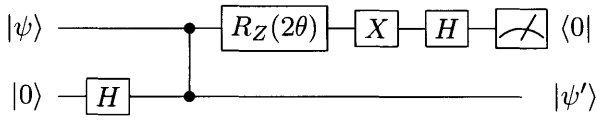
By commuting the $R_Z(\theta)$ rightwards on both lines, and noting that



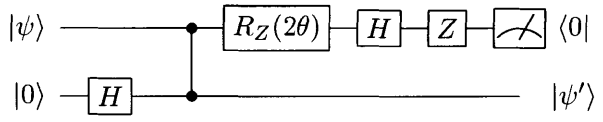
is equivalent to



we can see this performs the same quantum operation as



which since $HX = ZH$, is equivalent to



(5.28)

By direct computation, gadget (5.27) (which is equivalent to gadget (5.28)) performs the operation $HR_Z(2\theta)$. Let us call this gate $G_0(\theta)$. Likewise, if one postselects on $|1\rangle$, one obtains the gate $G_1(\theta) = XHR_Z(2\theta)$. (This gadget is well-known in MBQC; see below).

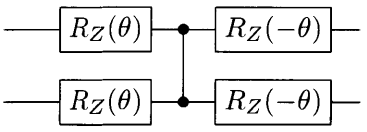
Therefore, by applying our gadgets (5.26) and (5.27), we can create postselected CCCs to simulate the evolution of a one-qubit circuit composed of gates in the set

$$\{H, XH, G_1(\theta), G_2(\theta)\}$$

. Intuitively, as long as the choice of θ is not pathological, these gates will generate all one-qubit gates. Therefore we have all one-qubit gates at our disposal via these gadgets. We will prove this statement rigorously in Lemma 5.6.5, which we defer to the end of this appendix. In fact, we show that as long as θ is not set to $k\pi/4$ for some integer k, then the set of one qubit gates generated by these gadgets is universal on a qubit. Thus postselected CCC's (where $\theta \neq k\pi/4$) can simulate arbitrary one-qubit operations.

To prove that postselected CCC's can perform universal quantum computation, we need to show how to perform an entangling two qubit gate. We can then appeal to the result of Brylinski & Brylinski [79] and Bremner et al. [71] that any entangling two-qubit gate, plus the set of all-one-qubit gates, is universal for quantum computation. But performing entangling two-qubit gates is trivial in our setup, since the Clifford group (and the conjugated Clifford group) contains entangling two-qubit gates. For example, we can easily perform the

controlled-Z gate between qubits with the following gadget:



(5.29)

This gadget clearly has the correct form, and hence composes with the gadgets (5.26) and 5.27 to form universal quantum circuits. This shows how to simulate BQP with postselected CCCs.

We can now recast this proof in the language of Measurement-Based Quantum Computing. Our result essentially follows from that fact that measuring graph states in the bases $HR_Z(2\theta)$ and H, combined with postselection, is universal for quantum computing. More formally, let E be series of Controlled-Z operations that create a graph state out of $H^{\otimes n}|0\rangle^{\otimes n}$ (we will specify the cluster state later). Let $U = R_Z(\theta)H$ for some θ to be specified later. Then consider creating the CCC for the Clifford circuit $C = X^S E$, where the notation X^S denotes that we apply an X gate to some subset $S \subseteq [n]$ of the qubits. We have that

$$\begin{aligned}
&H^{\otimes n} R_Z(-\theta)^{\otimes n} X^S E R_Z(\theta)^{\otimes n} H^{\otimes n}|0\rangle^{\otimes n} \\
=&H^{\otimes n} R_Z(-\theta)^{\otimes n} X^S R_Z(\theta)^{\otimes n} E H^{\otimes n}|0\rangle^{\otimes n} \\
=&H^{\otimes n} \left(X R_Z(2\theta)\right)^S E H^{\otimes n}|0\rangle^{\otimes n} \\
=&H^{\otimes n} \left(X R_Z(2\theta)\right)^S |\text{Cluster}\rangle \\
=&\left(\left(Z H R_Z(2\theta)\right)^S \otimes H^{\bar{S}}\right)|\text{Cluster}\rangle
\end{aligned}$$

Where the first equality follows from the fact that R_Z and E commute as they are both diagonal in the Z basis, the second follows from the fact that on the lines without an X the $R_Z(\theta)$ and the $R_Z(-\theta)$ cancel, and on the lines with an X we have $R_Z(-\theta)XR_Z(\theta) = XR_Z(2\theta)$, the third follows from the fact that E is constructed such that $EH^{\otimes n}|0\rangle^{\otimes n} = |\text{Cluster}\rangle$, and the fourth from the fact that $HX = ZH$. Now since we're measuring in the Z basis at the end of the circuit, the last row of Z's can be ignored, so the circuit is equivalent to:

$$\left(\left(H R_Z(2\theta)\right)^S \otimes H^{\bar{S}}\right)|\text{Cluster}\rangle$$

Now we simply need to show that measurement based quantum computation with postselection on such a state is universal for quantum computing. In other words, we need to show that if we can construct a Cluster state and measure some qubits in the H basis and others in the $HR_Z(2\theta)$ basis, and postselect on the outcomes, then we can perform universal quantum computation. It was previously known to be universal for MBQC if different θ's occur on each qubit [75]. In our setup we do not have this flexibility, but we instead have the additional ability to postselect.

Universality of this model follows from the fact that by preparing an appropriate Cluster state (using the standard trick to perform 1-qubit gates with MBQC), this gives us the ability to apply the one-qubit gate $HR_Z(2\theta)$ using postselection. Likewise, postselecting on $|1\rangle$ performs the operation $XHR_Z(2\theta)$. As discussed previously, by Lemma 5.6.5, as long as θ is not set to $k\pi/4$ for some integer k, this is a universal gate set on a qubit. The addition of entangling two-qubit operations on the Cluster state (namely, controlled-Z) boosts this model to universality.

□

We have now shown that postselected CCCs can perform BQP under postselection. We now extend this to show they can perform PostBQP = PP under postselection. This requires using the inverse-free Solovay-Kitaev algorithm of [212]. From this, the hardness result follows via known techniques [72, 11].

Theorem 5.6.3. *Postselected CCCs with $U = R_Z(\theta)H$ can decide any language in* PostBQP = PP, *for any choice of θ other than integer multiples of $\pi/8$.*

Proof. To prove this, we will apply Aaronson's result that Postselected BQP circuits, denoted PostBQP, can decide any language in PP. Aaronson's proof works by showing that a particular universal quantum gate set - namely the gate set G consisting of Toffoli, controlled-Hadamard, and one qubit gates - can decide PP under postselection.

We previously showed that our postselected CCCs can perform a different universal quantum gate G' consisting of controlled-Z, $HR_Z(2\theta)$, $XHR_Z(2\theta)$, H and XH. Therefore, in order to show that postselected CCCs can compute PP, we need to show how to simulate Aaronson's gate set G using our gate set $G'/$

One difficulty is that we must be extremely accurate in our simulation of these gates. This is because postselected quantum circuits may postselect on exponentially tiny events. Therefore, in order to simulate Aaronson's postselected circuits for PP, we will need to simulate each gate to inverse exponential accuracy.

Normally in quantum computing this simulation is handled by the Solovay-Kitaev Theorem, which roughly states that any universal gate set can simulate any other universal gate set to error ε with only polylog$(1/\varepsilon)$ overhead. Therefore with polynomial overhead, one can obtain inverse exponential accuracy in the simulation. This is why the choice of gate set is irrelevant in the definition of PostBQP. One catch, however, is that the Solovay-Kitaev theorem requires that the gate set is closed under inversion, i.e. for any gate $g \in G$, we have $g^{-1} \in G$ as well. This is an essential part of the construction of this theorem (which makes use of group commutators). It is an open problem to remove this requirement [99, 176]. As a corollary, it is open whether or not the class PostBQP can still compute all languages in PP if the gate set used is not closed under inversion. It is possible the class could be weaker with non-inversion-closed gate sets.

Unfortunately, the gate set G' we have at our disposal is not closed under inversion. Furthermore, since we obtained the gates using postselection gadgets, it is not clear how to generate the inverses of the gadgets, as postselection is a non-reversible operation. Therefore we cannot appeal to the Solovay-Kitaev theorem to show we can compute languages in PP.

Fortunately, however, even though our gate set does not have inverses, it does have a special property - namely, our set of one qubit gates contains the Pauli group. It turns out that recently, [212] proved a Solovay-Kitaev theorem for any set of one qubit gates containing the Paulis, but which is not necessarily closed under inversion. Therefore, by this result, even though our gate set is not closed under inversion, we can still apply any one-qubit gate to inverse exponential accuracy with merely polynomial overhead. So we can apply arbitrary one-qubit gates

It turns out this is sufficient to apply gates from Aaronson's gate set G consisting of Toffoli, controlled-H and one qubit gates with inverse exponential accuracy. To see this, first not that it is well-known one can construct controlled-V operations for arbitrary one-qubit gates V using a finite circuit of controlled-NOT and one-qubit gates - see [199] for details. Furthermore, it is possible to construct Toffoli using a finite circuit of one qubit gates and controlled-V operations [199]. This, together with the fact that controlled-NOT is equal to controlled-Z conjugated by Hadamard on one qubit, shows that each gate in G has an exact

decomposition as a finite number of controlled-Z gates and one-qubit gates. Hence, using controlled-Z gates and one-qubit gates compiled to exponential accuracy, one can obtain circuits from G with inverse exponential accuracy. Thus, our gate set G' can efficiently simulate gates from G, and hence our postselected CCCs can compute all languages in $\mathsf{PostBQP} = \mathsf{PP}$ as well.

□

From this, the hardness result follows via known techniques [72, 11].

Corollary 5.6.4. *Conjugated Clifford Circuits cannot be weakly simulated classically to multiplicative error unless the polynomial hierarchy collapses to the third level, for the choice of $U = R_Z(\theta)H$ for any θ which is not an integer multiple of $\pi/4$.*

To complete our proof, we merely need to show the following Lemma:

Lemma 5.6.5. *So long as θ is not an integer multiple of $\pi/4$, the gates $HR_Z(2\theta)$ and $XHR_Z(2\theta)$ are universal on a qubit. Furthermore, so long as θ is not an integer multiple of $\pi/4$, at least one of these gates is a rotation of the Bloch sphere by an irrational multiple of π.*

Proof. For convenience of notation, define $G_0 = HR_Z(2\theta)$ and $G_1 = XHR_Z(2\theta)$. We will actually begin by proving something stronger: namely, that as long as θ is not an integer multiple of $\pi/4$, then one of the rotations G_0 and G_1 is by an irrational multiple of π.

We will prove this by contradiction. Suppose that both G_0 and G_1 are rotations by rational multiples of π, call their rotation angles ϕ_0 and ϕ_1, respectively. By direct computation, first eigenvalue of G_0 is given by

$$\frac{1}{2\sqrt{2}}\left(-2\sin(\theta) - i\sqrt{6 + 2\cos(2\theta)}\right)$$

Since this must be equal to $e^{\pm\phi_0/2}$, and by considering the real part of this equation, we have that

$$\cos(\phi_0/2) = -\frac{\sin(\theta)}{\sqrt{2}} \tag{5.30}$$

By an identical argument, for gate G_1 we have that

$$\cos(\phi_1/2) = -\frac{\cos(\theta)}{\sqrt{2}} \tag{5.31}$$

Squaring these terms and summing them, we obtain that

$$\cos^2(\phi_0/2) + \cos^2(\phi_1/2) = \frac{1}{2}$$

Or, applying the fact $\cos^2 t = \dfrac{1 + \cos 2t}{2}$ and simplifying, one can see this is equivalent to

$$\cos(\phi_0) + \cos(\phi_1) + \cos(0) = 0$$

Since we are assuming by way of contradiction that G_0, G_1 are of finite order, we are assuming that ϕ_0, ϕ_1 are rational multiples of π. Previously, Crosby [96] and Włodarski [243] classified all possible solutions to the equation $\cos(\alpha_1) + \cos(\alpha_2) + \cos(\alpha_3)$ where each α_i are rational

multiples of π. The four possible solution families to this equation (assuming without loss of generality that $0 \leq \alpha_i \leq \pi$) are [243]

- $\{\beta, \pi - \beta, \pi/2\}$ where $0 \leq \beta \leq \pi$
- $\left\{\delta, \frac{2\pi}{3} - \delta, \frac{2\pi}{3} + \delta\right\}$ where $0 \leq \delta \leq \frac{\pi}{3}$
- $\left\{\frac{2\pi}{5}, \frac{4\pi}{5}, \frac{\pi}{3}\right\}$
- $\left\{\frac{\pi}{5}, \frac{3\pi}{5}, \frac{2\pi}{3}\right\}$

Since we have that one of our three angles is 0, the latter two cases are immediately ruled out, and we must have that the angles $\{\phi_0, \phi_1\}$ are either $\{\pi/2, \pi\}$ or $\{2\pi/3, 2\pi/3\}$. One can easily see that the first solution corresponds to $\theta = k\pi/2$ for an integer k, and the second solution corresponds to $\theta = k\pi/4$ for an odd integer k.

Therefore, so long as θ is not an integer multiple of $\pi/4$, we have a contradiction, as there are no further solutions to these equations where the ϕ_i are rational multiples of π. So if θ is set to any value other than $k\pi/4$ for an integer k , we have that at least one of the gates G_0 and G_1 is a rotation by an irrational multiple of π.

Now what remains to be shown is that the gates G_0 and G_1 are universal in the general case. This can be shown easily by the classification of continuous subgroups of $SU(2)$. The continuous subgroups of $SU(2)$ are $U(1)$ (corresponding to all rotations about one axis), $U(1) \times \mathbb{Z}_2$ (corresponding to all rotations about an axis a, plus a rotation by π through another axis perpendicular to a), and $SU(2)$. By our prior result we know that either G_0 or G_1 generates all rotations about its axis of rotation on the Bloch sphere. Therefore, if we can show that neither G_0 nor G_1 are rotations by angle π we are done, as these then must generate all of $SU(2)$. However this follows immediately from equations 5.30 and 5.31, since these equations imply that we can have either $\phi_0 = \pi$ or $\phi_2 = \pi$ only when θ is a rational multiple of $\pi/2$. Hence, as long as θ is not a rational multiple of $\pi/4$, neither G_0 nor G_1 is a rotation by π, and furthermore one is a rotation by an irrational multiple of π. These gates generate a continuous group which is neither $U(1)$ nor $U(1) \times \mathbb{Z}_2$, and therefore by the above observation these generate all of $SU(2)$.

□

5.7 Open Problems

Our work leaves open a number of open problems.

- What is the computational complexity of commuting CCCs? In other words, can the gate set CZ, S conjugated by a one-qubit gate U ever give rise to quantum advantage? Note that this does not follow from Bremner, Jozsa and Shepherd's results [72], as their hardness proof uses the gate set CZ, T or CCZ, CZ, Z conjugated by one-qubit gates. If this is true, it would say that the "intersection" of CCCs and IQP remains computationally hard. One can also consider the computational power of arbitrary fragments of the Clifford group, which were classified in [137]. Perhaps by studying such fragments of the Clifford group one could achieve hardness with lower depth circuits (see additional question below).

- We showed that Clifford circuits conjugated by tensor-product unitaries are difficult to simulate classically. A natural extension of this question is: suppose your gate set consists of all two-qubit Clifford gates, conjugated by a unitary U which is *not* a tensor product of the same one-qubit gate. Can one show that all such circuits are difficult to simulate classically (say exactly)? Such a theorem could be a useful step towards classifying the power of all two-qubit gate sets, i.e. resolving Conjecture 2.2.1.

- Generic Clifford circuits have a depth which is linear in the number of qubits [16]. In particular the lowest-depth decomposition for a generic Clifford circuit over n qubits to date has depth $14n - 4$ [188]. Such depth will be difficult to achieve in near-term quantum devices without error-correction. As a result, others have considered quantum supremacy experiments with lower-depth circuits. For instance, Bremner, Shepherd and Montanaro showed advantage for a restricted version of IQP circuits with depth $O(\log n)$ [73] with long-range gates (which becomes depth $O(n^{1/2} \log n)$ if one uses SWAP gates to simulate long-range gates using local operations on a square lattice). We leave open the problem of determining if quantum advantage can be achieved with CCCs of lower depth (say $O(n^{1/2})$ or $O(n^{1/3})$) with local gates only.

Chapter 6

Ball Permutations

In this chapter, we consider a completely different type of model "below BQP." Rather than considering the power of quantum computing with restricted gate sets, we will define a new model of quantum computing over a completely different Hilbert space, related to the scattering of distinguishable particles on a line. We call this the "Ball Permuting" model. We find surprising connections between this model and "one clean qubit model" of Knill and LaFlamme, despite the fact they are defined very differently. Furthermore, we show that the model is "encoded-universal" on certain input states.

This chapter is based on joint work with Scott Aaronson, Greg Kuperberg, and Saeed Mehraban [14].

6.1 Introduction

The standard model of quantum computing is defined using quantum circuits acting on qubits. The computational power of this model is captured by the complexity class BQP, which resides somewhere between the complexity classes BPP and PP [119]. However, physical systems can exist in Hilbert spaces which are not described by tensor products of qubits. For example, systems of noninteracting fermions, noninteracting bosons, or anyons in a 2+1 dimensional quantum field theory all live in Hilbert spaces with different mathematical descriptions.

A natural problem is to explore the computational complexity of these alternative physical systems. There are several reasons to study this problem. First, many of these alternative models of quantum computing seem to be intermediate in power between classical and quantum computing [159, 11, 195, 72], as discussed in Section 2.2.4. Therefore it is interesting to study their power from a purely complexity-theoretic standpoint, as they help delineate the boundary between classical and quantum computation. Second, these models sometimes have special properties from the perspective of mathematical physics. For instance, they might be "solvable" or "integrable" systems, which are regarded as simple to mathematical physicists. It is interesting to compare notions of simplicity in mathematical physics (solvability and integrability) to the notion of simplicity in computational complexity (efficient classical simulability).

Motivated by the above, in this chapter we consider an alternative model of quantum computing based on permuting quantum balls, and study its computational complexity. More specifically, we consider a system of n distinguishable particles ("balls") on a line. The computational basis of this Hilbert space consists of all permutations of the particles. We

denote this Hilbert space by $\mathbb{C}S_n$. The quantum operations in this model act on two balls at a time, and map the state $|x,y\rangle$ to the state $|x,y\rangle \rightarrow c|x,y\rangle + is|y,x\rangle$, where x and y distinct integers from 1 to n, and c and s are real numbers with $c^2 + s^2 = 1$. For example, if one has the state $|123\rangle$, and applies the above operation to the first two particles, the resulting state would be $c|123\rangle + is|213\rangle$. This is a quantum analog of exchanging the particles in that location with probability s^2; hence we call this the "partial swap" gate [1]. Physically, this "ball-permuting" model captures the scattering problem of distinguishable particles on a line with short-range interactions[2].

We study the computational complexity of several models based on the above formalism. First, we consider a model in which one starts in the state $|123\ldots n\rangle$, and then applies polynomially many partial swap gates. We show that one can approximate amplitudes in this model, within $1/\text{poly}$ additive error, within the one-clean-qubit model of Knill and LaFlamme, also known as the complexity class DQC1 [171]. This class captures the power of quantum computers in which all input qubits are in the maximally mixed state (i.e. a uniformly random basis state unknown to the experimenter) except one, which is in a pure initial state. This model is widely believed to be substantially weaker than BQP; in fact it is open whether or not DQC1 is even capable of universal *classical* computation. Therefore, the power of this model seems to be substantially weaker than BQP. Our result shows that exchange based quantum computing with distinguishable particles starting in the state $|12\ldots n\rangle$ yields a weak computational model. This in turn suggests that indistinguishability is a crucial computational resource for exchange based computations.

Next, we consider the computational power of this model when we begin with arbitrary initial states (i.e. states more complicated than $|1,2,\ldots,n\rangle$). First, we show that if the initial state is selected according to certain irreducible invariant subspaces, then this model can efficiently simulate BQP, using an encoded universality. We also mention an explicit construction based on the result that the exchange interaction on qubits is encoded-universal [103, 117]. Therefore allowing arbitrary initial states substantially boosts the power of this model.

Furthermore, we obtain a partial classification of the computational complexity of this model on different input states. In order to achieve this goal, we use the representation theory of the symmetric group. In particular, we use the Young-Yamanouchi orthonormal basis [158] to describe subspaces of our Hilbert space that are invariant and irreducible under the action of ball permuting gates. Therefore to understand the complexity of our model starting from an arbitrary input state, one merely needs to analyze its components in the Young-Yamanouchi basis using representation theory. We make progress towards this goal in Section 6.5.

One interesting finding of this classification is the discovery of a natural model of quantum computing which seems to be intermediate between DQC1 and BQP. In this model, we initially start with the input state

[1] Note Jordan [159] also uses permutations to obtain an intermediate model, but the operators he considers are different from partial swaps as they do not create superpositions in the particle location basis. Thus they are more analogous to "total swaps".

[2] In the interactions we consider, contact between particles is penalized with a delta function. In this sense the interactions are "hard". However, in the physics literature these are referred to as "soft short-range" interactions since they allow permutation of particles, whereas "hard" interactions forbid the permutation of particles.

$$\frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle|x\rangle,$$

consisting of n active qubits maximally entangled to n inert qubits. We imagine we can apply an arbitrary quantum circuit to the active qubits only (i.e. the left half of the state only), and then we measure both active and inert qubits together in the computational basis. If one were to only examine the computation on the left half of the state only, as well as the first qubit of the right half of the state, this model would be equivalent to DQC1 - because the left half of the state considered in isolation is maximally mixed. However, in this new model one additionally gets to observe the right half of the state. In other words, although the qubits are initially in an (unknown) uniformly random basis state at the start of the computation, one gets to learn what basis state they started in, but only *at the end of the computation.* We find that this model arises naturally in ball-permuting circuits starting with input states of specific "reducible subspaces". This model may be of independent interest.

In the full version of our paper [14], we consider two additional variants of the Ball Permuting Model. However we omit them in this thesis. The first is a restricted version of the model in which the exchange operators are required to satisfy the (parameter dependent) Yang-Baxter equation, which arises if one imagines that the balls have their own velocities on a line, and upon colliding, either exchange velocities (as in classical physics) or else pass through each other. We show that if intermediate measurements are added to this model, then one cannot sample from the same probability distribution (up to multiplicative error) efficiently with a classical computer unless the polynomial hierarchy collapses to the third level. This result might be somewhat surprising to mathematical physicists because systems obeying the Yang-Baxter equation in 1+1 dimensions are considered simple because they are "Bethe ansatz solvable" and "integrable". We also consider classical versions of this model, where a base AC^0 machine can query a deterministic, probabilistic, or non-deterministic ball-permuting oracle. We refer the interested reader to [14] for details of these results.

6.2 Models and Motivations

In this section we define the quantum ball-permuting model as a model of quantum computation, and then briefly explain how it models actual physical processes. The basic operations of this model are quantum swaps, as we will define them shortly.

The computational basis states in the ball permuting model are all $n!$ possible permutations on an n-element set (i.e., S_n). We start out in the state

$$|1, 2, \ldots, n\rangle.$$

At each time step, we get to pick any adjacent[3] pair of the n registers in our quantum state, and then apply an $(n^2 - n) \times (n^2 - n)$ unitary transformation that, for every pair of distinct labels $x \neq y$, maps

$$|x, y\rangle \rightarrow c|x, y\rangle + is|y, x\rangle,$$

[3]We get the same model if we allow swaps between any pairs. This is because we can simulate general swaps with adjacent ones.

where c and s are any two real numbers[4] satisfying $c^2 + s^2 = 1$. (We get to pick whichever c and s we like for each gate operation. However, c and s can't vary depending on the labels x and y. If one does allow for c and s to depend on x and y, then in Section 6.7 we show one recovers BQP, so this model is no longer a "weak model" of quantum computing).

Finally, we measure all n registers in the computational basis, and feed them to a classical computer for postprocessing.

We can represent the computational basis by the kets $|\sigma\rangle$ for any permutation $\sigma \in S_n$. We denote this Hilbert space by $\mathbb{C}S_n$. A quantum swap between t and $t+1$'th registers depends on one real parameter θ and is represented by the operator

$$X(\theta, t) = \cos\theta I + i \sin\theta L_{(t,t+1)}.$$

Here I is the identity operator, and L_σ is a representation of $\mathbb{C}S_n$ with the action[5]

$$L_\sigma|\tau\rangle = |\sigma \circ \tau\rangle.$$

The idea of this model is to capture n distinguishable particles moving around on a line (i.e., in $1+1$ dimensions). The only state that we care about is the order of the particles in the line. The x'th register of the quantum computer stores the label of the x'th particle, if the particles are listed in order from left to right. Whenever two particles meet, one of two things can happen: the particles can reflect, or they can pass through each other. The first happens with amplitude c, while the second happens with amplitude is.

The Yang-Baxter version of this model considered in the full version of this work [14] directly corresponds to the physics of particles interacting on a line. Here the interactions of the particles are constrained by both their velocities and the Yang-Baxter equation. In short, the model considered in this chapter is much simpler to define and work with, while the Yang-Baxter model is more directly related to the physics of interacting particles on a line.

6.3 The quantum ball permuting model

Based on the model introduced in Section 6.2, we will now formally define quantum ball-permuting complexity classes. We analyze their power with standard and arbitrary initial states in Sections 6.4 and 6.5, respectively. These sections assume that one can set the amplitudes c and s to arbitrary values for each interaction.

Definition 6.3.1. QBall *is the class of languages $L \subseteq \{0,1\}^\star$ for which there exist polynomial time Turing machines M and N such that on any input $x \in \{0,1\}^\star$, M outputs the description of a ball permuting quantum circuit C consisting of* $\mathrm{poly}(n)$ *partial swap gates such that, if N is run on inputs consisting of* $\mathrm{poly}(n)$ *samples from C in the computational basis*

- *If $x \in L$ then N accepts with probability at least* $\frac{1}{2} + \frac{1}{\mathrm{poly}(n)}$

[4]Note that up to a global phase this is the most general form of amplitudes we can consider which results in a *unitary* partial swap gate.

[5]The operator L acts from left to right on the basis of the Hilbert space. We can also talk about right multiplications $R(\tau)$ which map $|\sigma\rangle \mapsto |\sigma \circ \tau^{-1}\rangle$ for permutations σ and τ. The importance of these operators become clear in Section 6.5.2

- *Otherwise if $x \notin L$ then N accepts with probability at most* $\frac{1}{2} - \frac{1}{\text{poly}(n)}$.

The exact power of the class QBall each may depend on the input states allowed. In section 6.4 we analyze the power of the model when it start from a basis state and in section 6.5 we analyze it for arbitrary initial states.

6.4 Standard Initial States

In this section we will consider the power of QBall when the initial state is the identity permutation $|12\ldots n\rangle$. We observe the following containments for the ball permuting complexity classes

Theorem 6.4.1. QBall $\subseteq$ BQP.

To see this note that we can represent labels with binary strings using $O(\log n)$ qubits. By the Solovay-Kitaev theorem [99], any unitary on $O(\log n)$ qubits can be implemented using a polynomial-size quantum circuit, and therefore a BQP circuit can simulate a QBall circuit. Therefore the power of QBall is upper bounded by BQP; this model is no more powerful than standard quantum computing.

We will now show that starting from the initial state $|1, 2, \ldots, n\rangle$ the power of QBall is likely much weaker than that of BQP, because one can efficiently estimate individual amplitudes in QBall to $\frac{1}{\text{poly}}$ error in the complexity class DQC1, i.e. the class of quantum computations that can be performed with one clean qubit and $n-1$ maximally mixed qubits. DQC1 is believed to be a substantially weaker complexity class than BQP. (For a discussion of DQC1 see Section 2.2.4.) In contrast, for circuits over qubits it is BQP-hard to compute individual amplitudes to $\frac{1}{\text{poly}}$ accuracy. The following observation is crucial for the establishment of our result relating QBall and DQC1:

Lemma 6.4.1. *If C is any composition of X ball permuting operators over $\mathbb{C}S_n$, then $C|123\ldots n\rangle = |123\ldots n\rangle$ if and only if $C = I$.*

Proof. Suppose that $C|123\ldots n\rangle = |123\ldots n\rangle$, then permuting the labels arbitrarily gives $C|\sigma\rangle = |\sigma\rangle$ for all permutations $\sigma \in S_n$. □

The main result is the following:

Theorem 6.4.2. *There is a* DQC1 *algorithm which takes as input a description of a ball permuting circuit C over $\mathbb{C}S_n$, and which outputs a complex number α such that $|\alpha - \langle 123\ldots n|C|123\ldots n\rangle| \leq \frac{1}{\text{poly}(n)}$, with high probability* [6].

Proof. (Sketch) For any ball-permuting circuit, we know that for any two permutations π and π', $\langle\pi|C|\pi\rangle = \langle\pi'|C|\pi'\rangle$. This implies that $\langle 123\ldots n|C|123\ldots n\rangle = \frac{\text{Tr}(C)}{n!}$. On the other hand, Knill and Laflamme [171] showed for any quantum circuit U on n qubits composed of polynomially many gates, a DQC1 circuit can output a bit which is 1 with probability

[6] Furthermore, the DQC1 algorithm is able to find additive approximations for both the real and and imaginary values of the amplitude, separately.

$\frac{1}{2}+\left|\frac{\mathrm{Tr}(U)}{2^{n+1}}\right|$, and therefore estimate $|\mathrm{Tr}(U)/2^n|$ to $1/\mathrm{poly}$ additive error with postprocessing. Therefore it suffices to create a qubit unitary U on n qubits with the same trace as C on n particles, where 2^{n+1} is approximately the same as $n!$. Fortunately it is possible to do this using a carefully chosen encoding of permutations with qubits. In particular, we use an encoding of permutations that is both compressed and local. By compressed we mean that it uses $\log(n!\,\mathrm{poly}(n))$ bits, and by local we mean that we can use polynomial-size quantum circuits to simulate each quantum swap. We defer a proof of this fact to Section 6.8, where we discuss this issue in detail. □

A similar argument holds for arbitrary amplitudes $\langle\sigma|C|\sigma'\rangle$. This is followed by the reduction $C \mapsto L_\sigma C L_{\sigma'^{-1}}$ to the above problem. Therefore DQC1 can efficiently estimate amplitudes in the QBall model to $1/\mathrm{poly}$ accuracy. In contrast, for qubits, computing individual amplitudes to $1/\mathrm{poly}$ accuracy is BQP-hard. Therefore this is evidence that QBall is a weaker model of computation than BQP.

Note however Theorem 6.4.2 does not imply QBall $\subseteq$ DQC1 as decision languages, because we have only shown how to compute individual amplitudes of QBall in DQC1, while the decision class QBall may accept or reject based on an exponential sum of amplitudes. Likewise, Theorem 6.4.2 does not immediately imply a DQC1 machine could sample from the output distribution of a QBall circuit. Whether it is possible to efficiently sample the output distributions of ball permuting circuits with DQC1 computation is unknown.

We conjecture that this result can be generalized to a wide variety of quantum models based on group algebras. More precisely, consider a group G, with identity element e. Then construct the Hilbert space $\mathcal{H}_G$ with orthonormal basis $\{|g\rangle : g \in G\}$. Let $\mathbb{C}G$ be the (left) group algebra, and $x \in \mathbb{C}G$. Then $\langle e|x|e\rangle = \frac{1}{|G|}\mathrm{Tr}(x)$, which is a reduction to the computation of normalized trace. If one has a compressed and local encoding of G as described in the proof of Theorem 6.4.2, then this would imply that amplitudes of computations over $\mathbb{C}G$ can be simulated in DQC1 as well. We leave this is an open problem.

6.5 Arbitrary Initial States

We already observed that the model we obtain seems to be restricted if one starts with and measures according to the computational basis. In this section, we further examine how the power of ball-permuting model depends on the input states. We first give a simple construction showing that QBall is universal for BQP when given particular input states. The proof is based on DiVincenzo *et al.*'s result that the exchange interaction is universal for quantum computing [103]. We then provide a partial classification of the power of QBall on different input states using the representation theory of S_n. This requires substantial work in representation theory, and is the main technical contribution of this work. In Section 6.5.2 we describe a number of input states which boost the power of QBall up to BQP. We then describe some other input states which yield a model intermediate between DQC1 and BQP in Section 6.6.

6.5.1 A simple proof that QBall = BQP on arbitrary initial states

First of all, building on DiVincenzo *et al.*'s result that the exchange interaction is universal for quantum computing [103], we observe that when the initial state need not be a computational basis state, the quantum ball permuting model has the full power of BQP. The

proof uses the notion of *encoded universality*: although our model does not allow arbitrary unitaries on the Hilbert space $\mathbb{C}S_n$, it can simulate arbitrary unitaries on certain subspaces of $\mathbb{C}S_n$ which encode qubits. We can therefore perform universal quantum computation on the encoded subspace, assuming our inputs are allowed to lie in the encoded subspace. This can be summarized by the following theorem:

Theorem 6.5.1. *If* QBall *is allowed to have non-basis input states, then* QBall = BQP.

The authors of [103, 42] showed that one could use the exchange interaction to simulate one logical qubit using three physical qubits by the following encoding

$$|0_L\rangle := \frac{|010\rangle - |100\rangle}{\sqrt{2}}$$

and,

$$|1_L\rangle := \frac{|010\rangle + |100\rangle - 2|001\rangle}{\sqrt{6}}.$$

Moreover the author of [103] showed how to implement an approximate CNOT on logical qubits using exchange interactions. This result was further improved in [117], where the authors found closed form expressions for this implementation.

In our model, we mimic this encoded universality construction for qubits using permutations. Specifically, to encode a logical qubit, we use permutations of a three-element set. We let the permutation labels 1 and 2 represent the qubit state $|0\rangle$ and the permutation label 3 represent the qubit state $|1\rangle$. We then symmetrize over the labels which represent with zeros and over the labels which represent with ones, to obtain states over $\mathbb{C}S_3$ which represent each basis state $|001\rangle, |010\rangle, |100\rangle$ used in DiVincenzo et al.'s construction. For example we represent the qubit state $|001\rangle$ with the state $|123\rangle + |213\rangle$, and we represent the state $|010\rangle$ with the state $|132\rangle + |231\rangle$. To encode logical zero and logical 1, we use DiVincenzo *et al.*'s encoding, ported over to permutations using the above correspondence. One can check this simulation works because the only operators being used in both models are permutations. For $n > 3$ qubits we use the same symmetrization idea to simulate exchange interactions on $(\mathbb{C}^2)^{\otimes n}$; a detailed proof of this fact in given in Sections 6.9.1 and 6.9.2.

6.5.2 Partial classification of input states which make QBall = BQP

In this section we demonstrate a partial classification for the computational power of this model according to different initial states. Our objective is to demonstrate that different input states lead to different interesting models of computation. This classification is obtained using the representation theory of the symmetric group. (For a brief review of this theory see Section 2.3.2).

A representation of a group G is a homomorphism : $G \to GL(V)$, for some vector space V, which obeys the same multiplication rule as the group law. We interchangeably refer to V or the homomorphism itself as the representation. The regular representation of S_n with left action is according to the homomorphism $L : S_n \to GL(\mathbb{C}S_n)$, with the map $L_g : |h\rangle \mapsto |g.h\rangle$. Similarly, the right action $R : S_n \to GL(\mathbb{C}S_n)$, is according to the map $R_g : |h\rangle \mapsto |h.g^{-1}\rangle$. An invariant subspace is a subspace that is stable under the action of a particular representation, i.e., the image of this subspace under the action of the group is equal to the subspace itself. A representation is called irreducible if its only invariant subspaces are the singleton $\{0\}$ and the representation itself.

Under these (left and right) regular representations the Hilbert space $\mathbb{C}S_n$ decomposes into irreducible representations as

$$\mathbb{C}S_n \cong \bigoplus_{\lambda \vdash n} V_\lambda \otimes X_\lambda,$$

where each λ is a partition of the number n: that is, a list of non-negative integers in non-ascending order summing to n. $\lambda \vdash n$ means that λ is a partition of n. $V_\lambda \otimes X_\lambda$ is a summand of the decomposition the tensor product of two vector spaces with $\dim(V_\lambda) = \dim(X_\lambda) =: d_\lambda$ and $\sum_\lambda d_\lambda^2 = n!$.. Interestingly, the left actions of S_n only acts on V_λ's and act trivially on the X_λ's. X_λ's are called the multiplicity spaces corresponding to each V_λ, as they enumerate all the irreducible representations isomorphic to V_λ.

Denote the Lie group generated by ball-permuting operators with G_n, and the projection of G_n onto V_λ with $G_n(V_\lambda)$. Note that $G_n(V_\lambda)$ is unitary, as it corresponds to the action of G_n on the subspace of the irrep V_λ.

Theorem 6.5.2. *If $\lambda \vdash n$ or its transposed partition consists of two parts, then $SU(V_\lambda) \subseteq G_n(V_\lambda)$.*

Proof. (Sketch) By induction, we first prove the statement for $n = 3$. We then use subgroup adapted Young-Yamanouchi basis which manifests the branching rule (discussed in Section 2.3.2), along with decoupling lemma and bridge lemma of Aharonov and Arad [28] to deduce the statement of this theorem for each subspace one-by-one. This procedure fails for subspaces corresponding to general partitions with more than two rows (columns). □

The transpose of a partition λ is a partition whose rows are the columns of λ and whose columns are the rows of λ.

Therefore, if the input state to QBall corresponds to an irreducible representation which a) consists of two parts and b) is sufficiently large in dimension, then by Theorem 6.5.2 one can perform encoded universal computation on this input state. The description of these input states is given in Section 6.9.3. We believe that the result can be extended to partitions with more than two rows or columns, however, the tools we used are restricted and we need more ideas to achieve this goal. We leave this result as an open question.

6.6 Some new intermediate quantum computing models

We saw in Section 6.4 that for a ball permuting circuit C, the amplitude $\langle 12\ldots n|C|12\ldots n\rangle$ can be additively approximated using a DQC1 algorithm. Moreover, the last Section asserted that for specific partitions λ, if $|\psi\rangle$ is a separable state over the partition $V_\lambda \otimes X_\lambda$, then it is BQP-complete to compute the amplitude $\langle\psi|C|\psi\rangle$ within additive error. In this Section we provide evidence that the ball-permuting model along specific subspaces of $\mathbb{C}S_n$ yields a model of computing that is intermediate between DQC1 and BQP.

Suppose that instead of the computational basis, we initialize the ball-permuting model with the projection of the state $|123\ldots n\rangle$ onto an irrep λ. Denote this (normalized) state by $|\lambda\rangle$. Then we apply a sequence of ball-permuting gates, and at the end of the computation we sample a pair of tableaux in $V_\lambda \otimes X_\lambda$ according to the Young-Yamanouchi basis (see Section 2.3.2 for a review). Inspired by this model, we formally define the following complexity class:

Definition 6.6.1. SampQBall(λ) *is the class of problems that are solvable in polynomial time using polynomially many samples from the above model.*

The exact power of this model is unknown, however, we motivate reductions to an interesting complexity class that is intermediate between DQC1 and BQP. We can also define the following computational problem:

Definition 6.6.2. *(QBall(λ)) given a ball permuting circuit C, and a partition λ, evaluate an additive approximation to $\langle\lambda|C|\lambda\rangle$.*

Interestingly, the initial state $|123\ldots n\rangle$ is equally supported on all of the irreps λ of the decomposition $\mathbb{C}S_n \cong \bigoplus_{\lambda\vdash n} V_\lambda \otimes X_\lambda$, and moreover it is maximally entangled over each partition $V_\lambda \otimes X_\lambda$. Also, notice that (as mentioned in the last Section) all ball-permuting operators only act on V_λ parts, and act trivially on X_λ's, the multiplicity spaces [7].

In short, we find that the model operates on one half of a maximally entangled state, while leaving the other half untouched. At the end of the computation, one measures both halves of the maximally entangled state. To investigate the power of this model, as a toy model, we consider a restricted model of quantum computing. Imagine we have access to only one half of the state

$$|\psi\rangle = \frac{1}{2^{n/2}} \sum_{x\in\{0,1\}^n} |x\rangle|x\rangle.$$

Suppose we can perform standard unitary quantum computation on the left half of the state, but one cannot access the right half of the state. At the end of the computation, then we get to measure both halves of the state in the computational basis.

This model seems very similar to the class DQC1, because if C is a quantum circuit on the n active qubits, then

$$\langle\psi|C|\psi\rangle = \frac{Tr(C)}{2^n},$$

the normalized trace. This is simply because the reduced density matrix of the left half of the state is maximally mixed. We have seen in Section 6.4 that evaluation of such a trace within additive error is complete for the class DQC1. Therefore if we define the trace computing class:

Definition 6.6.3. *(Trace computing quantum polynomial time)* TQP *is the class of problems that are polynomial-time reducible to additive approximation of $\langle\psi|C|\psi\rangle$, the normalized trace of the matrix. Also* SampTQP *is the class of problems that are solvable with high probability using polynomially many samples from $C|\psi\rangle$, in the computational basis.*

Then we trivially find TQP = DQC1. However, the class SampTQP seems to be more powerful than DQC1. The reason is that one gets to measure the right half of the state at the end of the computation; this is not an ability one has in DQC1. To put it another way, DQC1 is defined using maximally mixed states as inputs, which is equivalent to performing your computation on a random basis state $|x\rangle$. SampTQP is equivalent to performing your computation on a random $|x\rangle$, but at the end of the computation, you get to learn which $|x\rangle$ you started with. As a result, it appears that SampTQP is intermediate between DQC1 and BQP. Note that SampTQP is BQP-universal under postselection as well.

We first observe that:

[7] Intuitively, the situation resembles a quantum/classical hybrid memory in the sense of [174], where classical bits enumerate the name λ corresponding to the particular irrep, and the quantum memory corresponds to a bipartite system-environment Hilbert space $V_\lambda \otimes X_\lambda$; V_λ plays the role of a system, and X_λ is its environment which is inert and also maximally entangled with the system.

Theorem 6.6.1. *If λ is a partition with two equally sized rows, then* $\mathsf{QBall}(\lambda) \in \mathsf{TQP}$.

Proof. (Sketch) following the proof of theorem 6.4.2 we use a compressed and local representation of standard tableaux with two rows of length $n/2$ using strings of bits. Specifically, suppose we represent standard tableaux on two equal-sized rows by bit strings of length n. The ith entry represents whether the number i appears on the top or bottom row. There are 2^n such strings and $2^n/\operatorname{poly}(n)$ valid tableaux, so although many of the strings do not represent valid tableaux, a $1/\operatorname{poly}$ fraction do represent valid tableaux. So this encoding is very efficient. Furthermore, it is local in the sense that to exchange two labels, one simply exchanges their corresponding bits. Finally, it is easy to test if a string is a valid encoding. Additionally, it one can compute using $O(\log n)$ ancillae the axial distance (as defined in Section 2.3.2) between two labels in the tableau. Following the techniques of Section 6.8, the existence of such an encoding implies membership in DQC1, as one can simulate the action of ball-permuting gates on these basis, and use a DQC1 procedure to evaluate an additive approximation to $\langle\lambda|C|\lambda\rangle$ .See Section 6.8 for details.

□

We conjecture that the power of QBall on all such input states $|\lambda\rangle$ is contained in TQP. However, it is difficult to prove this fact because DQC1 is defined using qubits, while the $|\lambda\rangle$ basis is labeled by the Young-Yamanouchi basis (See Section 2.3.2). As in Section 6.4, in order to prove QBall is in TQP on this input states, we would need to find an encoding of the Young-Yamanouchi basis which is both extremely compressed and local for arbitrary Young diagrams. We mention this is an open problem in Section 6.10.

Next, we build a connection between the two classes $\mathsf{SampQBall}(\lambda)$ and SampTQP:

Theorem 6.6.2. *If λ is a partition with two equally sized rows,* $\mathsf{SampQBall}(\lambda) \subseteq \mathsf{SampTQP}$.

Proof. (Sketch) as in the proof of theorem 6.6.1, we use a compressed and local representation of tableaux with two equally sized rows using binary strings to simulate the action of ball permuting gates on V_λ in a succinct space, and sample from the output distribution in the computational basis. After postprocessing, with high probability the sampled string corresponds to a valid sample from $\mathsf{SampQBall}(\lambda)$. □

SampTQP is a restricted model of computation on qubits, and is an interesting model on its own right. We can immediately observe the following

Theorem 6.6.3. $\mathsf{DQC1} \subseteq \mathsf{SampTQP} \subseteq \mathsf{BQP}$.

Proof. $\mathsf{SampTQP} \subseteq \mathsf{BQP}$ is immediate. To see $\mathsf{DQC1} \subseteq \mathsf{SampTQP}$ do a DQC1 computation, assuming optimistically that the first active bit is in the pure state $|0\rangle$. Then, at the end, when we measure we will find out whether or not the assumption was correct, and it will have been with probability $1/2$. □

6.7 Label-Dependent Exchange Interactions Yield BQP

In the definition of $X(\cdot,\cdot)$ operators, the angle θ is independent of the labels that are being swapped. Here we consider the power of a different model in which θ depends on the labels being swapped, and show that this model has power equivalent to BQP.

Consider the local unitary, $Z(\tilde{\theta}, k)$, wherein the transposition angles depend on the color of the labels. Here $\tilde{\theta} = \{\theta_{ij}\}$ is a list of angles, one element per each $i \neq j \in [n]$. By definition $Z(\tilde{\theta}, k)$ acts on the labels $|ab\rangle$ in the locations k and $k+1$ with the following map

$$Z(\tilde{\theta},k)|ab\rangle = \cos\theta_{ab} I + i \sin\theta_{ab} L_{(a,a+1)}.$$

If we assume real valued angles with $\theta_{ij} = \theta_{ji}$, then the operator Z becomes unitary. Clearly, the X operators are the special case of the Z operators. In order to see this, consider any basis $|\sigma\rangle, \sigma \in S_n$, and suppose $\sigma(k) = a, \sigma(k+1) = b$ then

$$Z^{\dagger}(\tilde{\theta},k)Z(\tilde{\theta},k)|\sigma\rangle = (\cos\theta_{ab} - i\sin\theta_{ab} L_{(k,k+1)})(\cos\theta_{ab} + i\sin\theta_{ab} L_{(k,k+1)})|\sigma\rangle = |\sigma\rangle.$$

Next we use a simple encodings of qubits using labels $1, 2, 3 \ldots, n$, and the Z operators to operate on them as single and two qubit gates and prove that this modified model can simulate BQP. More specifically, we prove that using a sequence of Z operators, one can encode any element in the special orthogonal group. For an example of encoded universality see [135, 41]. We encode each qubit using two labels. Given two labels $a < b$ we define the encoded (logical) qubits as

$$|0\rangle := |ab\rangle$$

and,

$$|1\rangle := i|ba\rangle.$$

Using simple $X(\theta, 1)$ we can apply arbitrary rotation of the following form

$$|0\rangle \rightarrow \cos\theta|0\rangle + \sin\theta|1\rangle$$

and,

$$|1\rangle \rightarrow \cos\theta|1\rangle - \sin\theta|0\rangle.$$

We are dealing with orthogonal matrices which are represented over the field or real numbers. Using the Z operators, we can discuss a controlled swap of the form

$$S(i,j,k,l) := Z(\pi/2\delta_{i,j}, k, l).$$

In simple words, $S(i,j,k,l)$ applies the swap $iL_(k,l)$, on the k and l'th labels if and only if the content of these label locations are i and j (j and i). We can also extend it to the following form

$$S(\{(i_1,j_1)^{s_1}, (i_2,j_2)^{s_2}, \ldots, (i_t,j_t)^{s_t}\}, k, l) := Z(\pi/2\delta_{i,j}, k, l).$$

Where s_m can be a symbol $\star$ or nothing. Given $(i_m, j_m)^{\star}$ in the list means that the swap $(iL_{(k,l)})^{\dagger} = -iL_{(k,l)}$ is applied if the content of k and l are i_m and j_m. And given plain (i_m, j_m) in the list means $iL_{(k,l)}$ if the content of k and l are i_m and j_m.

Suppose that one encodes one qubit with labels $a < b$ and another one with $x < y$, we wish to find a unitary operator which applies a *controlled not* on the two qubits, that is the following map

$$
\begin{aligned}
|00\rangle &:= |a,b,x,y\rangle \rightarrow |a,b,x,y\rangle = |00\rangle \\
|01\rangle &:= i|a,b,y,x\rangle \rightarrow i|a,b,y,x\rangle = |01\rangle \\
|10\rangle &:= i|b,a,x,y\rangle \rightarrow -|b,a,y,x\rangle = |11\rangle \\
|11\rangle &:= -|b,a,y,x\rangle \rightarrow i|b,a,x,y\rangle = |10\rangle
\end{aligned}
$$

It can be confirmed that the following operator can do this

$$C := S(\{(a,x),(a,y)^\star\},1,3)S(\{(a,x),(a,y)\},2,3)S(\{(a,x),(a,y)\},1,2).$$

Given these two operators, one can simulate special orthogonal two-level systems, that is for each orthonormal $|\psi\rangle$ and $|\phi\rangle$ in the computational basis of n qubits we can apply an operator which acts as

$$|\psi\rangle \rightarrow \cos\theta|\psi\rangle + \sin\theta|\phi\rangle.$$

and,

$$|\phi\rangle \rightarrow \cos\theta|\phi\rangle - \sin\theta|\psi\rangle.$$

6.8 Detailed proofs for Section 6.4

Theorem 6.8.1. *There is an efficient* DQC1 *algorithm which takes the description of a* poly(n) *size ball permuting circuit C over $\mathbb{C}S_n$ as its input, and outputs a complex number α such that*

$$|\alpha - \langle 123\ldots n|C|123\ldots n\rangle| \leq \frac{1}{\mathsf{poly}(n)}$$

with high probability.

The theorem is proved in three steps. First, in lemma 6.8.1 we observe that for ball permuting circuits the computation of single amplitudes can be reduced to the computation of (normalized) traces. Next, we borrow a result of [222] which provides a reduction from additive approximation of traces for unitary matrices to DQC1 computations. Finally, in the third step, by some careful analysis it is shown that the DQC1 reduction of the second step is an efficient one. The main idea for this step is to use a compressed encoding of permutations with binary bits.

The amplitudes in ball permuting circuits are related to traces according to:

Lemma 6.8.1. *For any ball permuting quantum circuit C, the trace*

$$Tr(C) = n!\langle 123\ldots n|C|123\ldots n\rangle.$$

Proof. A quantum ball permuting circuit, by definition, consists of left permuting actions only which commute with right actions $R(\sigma)$ (relabeling) for any $\sigma \in S_n$. Thereby,

$$
\begin{aligned}
\langle 123\ldots n|C|123\ldots n\rangle &= \langle 123\ldots n|R^{-1}(\sigma)CR(\sigma)|123\ldots n\rangle \\
&= \langle\sigma|C|\sigma\rangle.
\end{aligned}
$$

From this, $Tr(C) = \sum_{\sigma \in S_n} \langle \sigma | C | \sigma \rangle = n! \langle 123 \ldots n | C | 123 \ldots n \rangle$. □

Next, we formally mention the problem of trace approximation:

Definition 6.8.1. *(*Trace*) given as input the* poly(n) *size description of a unitary circuit U as a composition of gates from a universal gate set over n qubits, compute a complex number t such that $|t - \frac{1}{2^n} Tr(U)| \leq \frac{1}{\text{poly}(n)}$, with high probability.*

The following theorem provides an efficient DQC1 algorithm for Trace:

Theorem 6.8.2. *(Jordan-Shor [222])* Trace *is a complete problem for* DQC1. [8]

Indeed, this theorem can be reformulated as: given an n qubit unitary U, there is a round of DQC1 computation which reveals a coin which gives heads with probability $\frac{1}{2} + \frac{1}{2} \frac{\Re Tr(U)}{2^n}$. Also, there is another similar computation which gives a coin with bias according to the imaginary part of the normalized trace.

Using these observations, we are ready to give a proof for the main theorem:

Proof. (of theorem 6.8.1) The objective is find an efficient algorithm which given a ball permuting circuit C over n labels, outputs the description of a unitary U over $m = \text{poly}(n)$ qubits such that $\frac{1}{2^m} Tr(U) = \frac{1}{\Delta(n)} \langle 123 \ldots n | C | 123 \ldots n \rangle$, with $\Delta(n) = \text{poly}(n)$. Given this reduction using theorem 6.8.2 we deduce that the additive approximation of the amplitude can be obtained by rounds of DQC1 computation.

The basic idea is to encode permutations with strings of bits, perform the circuit C on this encoded space, and take the trace of C using a DQC1 circuit. For ease of presentation, we will first present the proof using a simple encoding of permutations which turns out not to work, and later describe the more complex encoding which suffices for the proof.

Suppose we represent a permutation $\sigma \in S_n$ using $n\lceil \log n \rceil$ bits, i.e. each particle label in $[n]$ is represented using $\lceil \log n \rceil$ bits. Simulate each X gate in C with a quantum circuit which swaps the encoded numbers in a superposition. Since each gate only acts on $O(\log n)$ qubits, such a quantum circuit can be efficiently obtained from a universal gate set by the Solovay-Kitaev Theorem [99]. Let U be the composition of these unitary circuits. The objective is to perform a DQC1 computation to obtain an approximation to $Tr(U)/D$, where D is the dimension of the Hilbert space that U is acting on. However one can easily see that $\text{Tr}(U)$ is not in general equal to $\text{Tr}(C)$ - because among the summands of $\text{Tr}(U)$ there are terms like $\langle b | U | b \rangle$, where b is a string of bits with repeated labels (for example $|11234\rangle$). Such terms do not appear in $\text{Tr}(C)$ because they are not valid encodings of permutations. In order to avoid the contribution of these terms, use $\frac{n(n-1)}{2}$ more (flag register) qubits, $f_{ij}, i < j \in [n]$. Then we add another term T to the quantum circuit to obtain UT. The role of T is simply to modify the flag registers in a way that the contribution of unwanted terms in the trace becomes zero: for each $i < j \in [n]$, using sequences of $CNOT$ gates, T compares the qubits $(i-1)\lceil \log n \rceil + 1$ to $i\lceil \log n \rceil$ with the qubits $(j-1)\lceil \log n \rceil + 1$ to $j\lceil \log n \rceil$, bit by bit, and applies NOT to the register $f_{i,j}$ if the corresponding bits are all equal to each other. Then UT is fed into the Trace computation. Let's see what approximation

[8] Moreover, the authors show that Trace is a complete problem for this class, with polynomial time pre-processing.

to $\langle 123\ldots n|C|123\ldots n\rangle$ we get in this case. Let $N := n\lceil \log n\rceil + n(n-1)/2$. The trace $Tr(U) = \sum_{x\in\{0,1\}^N} \langle x|U|x\rangle$. Given the described construction, the term $\langle x|U|x\rangle = \langle\sigma|C|\sigma\rangle$, if and only if the label part of x is the correct encoding of the permutation σ, and if x is not a correct encoding of a permutation it gives 0. There are $2^{n(n-1)/2}$ strings like x which encode σ correctly, therefore

$$\frac{Tr(U)}{2^N} = \frac{2^{n(n-1)/2}}{2^N} Tr(C) = \frac{n!}{2^{n\lceil \log n\rceil}} \langle 123\ldots n|C|123\ldots n\rangle.$$

Using DQC1 computations we can estimate the value of $Tr(U)/2^N$ to 1/poly additive error. This is almost what we want, but the problem is that the coefficient $\frac{n!}{2^{n\lceil \log n\rceil}}$ can be exponentially small, because $\log(n!) \approx n\log n - \Theta(n)$ by Stirling's approximation. Therefore the amplitude we are trying to compute ($\langle 123\ldots n|C|123\ldots n\rangle$) is exponentially suppressed in this model, so a 1/poly approximation to $1/2^N Tr(U)$ does not yield a 1/poly approximation to $\langle 123\ldots n|C|123\ldots n\rangle$.

Taking a close look at the this coefficient, one can see that for any encoding of permutations with bit-strings, the proportionality constant appears as

$$\frac{n!}{dim V}$$

where V is the dimension of the Hilbert space that is used to encode permutations in it. In the latter example, we used $O(n\log n)$ bits to encode permutations of n labels. Our problem arose because $2^{n\log n}$ is exponentially larger than $n! \approx (n/e)^n$.

To fix this issue, we will need to use a more compressed encoding. More precisely, we need an encoding that uses $O(\log(n!\mathsf{poly}(n))$ bits. Moreover, in order to provide efficient quantum circuits, the code needs to be local, in the sense that in order to apply a swap, we just need to alter only $O(\log n)$ bits. Otherwise, it is not clear if it is possible to implement the encoded swaps efficiently with qubit quantum circuits.

To do this, we consider encoding permutations using $\lceil \log n!\rceil$ bits. Specifically, we consider an ordering of permutations (called the factorial number system, reviewed in Section 6.8.1) and represent each permutation using $\lceil \log n!\rceil$ bits. This encoding is extremely efficient, but it is not local, because one may need to rewrite all of the bits to perform an encoded swap. To overcome this issue, to perform an encoded swap on indices i and $i+1$, we first extract $O(\log n)$ bits of information which encodes the values of the permutation at those locations. We then apply the partial swap on the extracted entries (which is now manifestly local) and then convert the inefficient codes back to the compressed ones. In this manner we can approximate $\langle 123\ldots n|C|123\ldots n\rangle$ to 1/poly accuracy in the manner described above.

More precisely, we encode permutations using the factorial number system, described in the next Subsection 6.8.1. The basic idea is that once one has specified the first k entries of the permutation, there are only $n-k$ choices for the next entry. Therefore, one can specify the next entry of the permutation by indicating which of these remaining $n-k$ elements to choose. In particular, we can represent a permutation σ by a series of numbers $a_n, a_{n-1}, \ldots a_2$, where each a_i is a number from 0 to $i-1$ indicating which of the remaining items appears next in the permutation. (Note a_1 need not be included, since once you have specified the first $n-1$ entries of the permutation, you need not specify the last entry.) The

permutation is then represented by the number $N_\sigma = \sum_i a_i(i-1)!$, which ranges from 0 to $n! - 1$.

This representation of a permutation is manifestly local - in order to swap two entries i and $i+1$, one merely needs to perform some operation on a_i and a_{i+1}. (This operation is slightly more complicated than just switching a_i and a_{i+1} due to an edge case where a_i and a_{i+1} have adjacent labels amongst the remaining labels, but as explained in the next subsection this is still local). Furthermore, one can easily extract a_i in polynomial time from the number N_σ. To see this, if we let r_i be the remainder of $N_\sigma/i!$, then a_i is simply equal to quotient of $r_i/(i-1)!$. This is analogous to the fact that one can efficiently extract the ith digit base 10 of a number encoded in binary. Likewise, given new values of a_i and a_{i+1} one can easily update the value of N_σ to its new value.

One subtlety in this approach is that when extracting a_i and a_{i+1}, we are very restricted in our use of ancillae. In particular, since DQC1 circuits only have access to maximally mixed ancillae, we can only ever simulate the use of $O(\log n)$ pure ancillae. This is because, if we ensure the all zero string in the ancillae goes back to the all zero string at the end of the computation, then we can postselect them to be all 0's at the end of the DQC1 computation. Since this occurs with 1/poly probability this is within our abilities.

Therefore, in order to complete this argument, we will need to show that a_i and a_{i+1} can be extracted, and N_σ be updated after the swap, using only $O(\log n)$ pure ancillae. In fact, we will show one can get away with only $O(1)$ ancillae. This is because given an integer N, for a fixed[9] integer k, it is possible to compute the quotient q and remainder r of N/k using $O(1)$ ancillae. Indeed the grade-school long division algorithm suffices for this task, and uses only 2 ancillae (we thank Luke Schaeffer for pointing this out to us). Suppose one wishes to compute the quotient q and remainder r of $N/k = qk + r$. To do so, simply compute how many times k divides the first $\lceil \log k \rceil$ bits of N, store this as the first bit q_i of the quotient, and subtract $q_i 2^{\lceil \log N \rceil - \lceil \log k \rceil}$ from N. Repeat. One can easily see that since we're dividing in binary, for every bit we compute of the quotient (with the possible exception of the first bit), the leading bit of N is set to zero. Therefore we can reuse this space to store an additional bit of q. At the end of the computation one has q and r stored in $\lceil \log q \rceil + \lceil \log k \rceil \leq \log N + 2$ bits. We have therefore computed the quotient and remainder reversibly in place with only two ancillae. This suffices to prove one can extract a_i and a_{i+1} reversibly from N_σ using only $O(1)$ ancillae. The proof that one can update N_σ in-place reversibly after altering a_i and a_{i+1} follows analogously by running the above operation in reverse. Therefore this encoding of permutations can be used to estimate $\langle 123\ldots n|C|123\ldots n\rangle$ to 1/poly error in DQC1.

Note that the construction is based on adjacent swaps only. If in the description of C nonadjacent swaps are implemented, we can simulate these swaps by adjacent ones. We construct U by approximating each adjacent X gate in C. Each such gate alters $O(\log n)$ bits and because of the Solovay-Kitaev theorem [99], there exists a $\mathsf{poly}(n, \log 1/\varepsilon)$ size circuit that approximates each X gate within error ε. □

6.8.1 Factorial number system

The encoding of each permutation, $\sigma(1), \sigma(2), \ldots, \sigma(n)$ ($\sigma \in S_n$), is accomplished by a walk from root to each leaf of the following tree, T_n: consider a tree with its root located at node

[9] Since our swaps are specified ahead of time in the description of C, we can hard-code the numbers we divide by into the circuit.

0, as we mark it to be distinct. Let node 0 have degree n, with its children marked with numbers $1, 2, 3, \ldots, n$, from left to right. Denote these nodes by layer 1. Let each node of layer 1 have $n-1$ children, and label each child of node i in layer 1, by numbers $[n] - \{i\}$, in an increasing order from left to right. Construct the tree inductively, layer by layer: each node k in layer j have $n-j$ children, and the children labeled with numbers $[n] - L_k$. Where L_k is the set of labels located on the path from node 0 to node k. Therefore, nodes of layer n have no children. The number of leaves of the tree is $n!$. For each leaf there is a unique path from root down to the leaf, and the indices from top to down represent a permutation. This is because the indices of each path are different from each other. Also each permutation σ is mapped to a unique path in this tree: start from node 0, pick the child with index $\sigma(1)$, then among the children of $\sigma(1)$, pick the child with index $\sigma(2)$ and so on. Therefore, this establishes a one-to-one map between the paths on T_n and permutations of labels in $[n]$.

The next step is to provide a one-to-one mapping from the paths on the graph to bit strings of length $\log n! + O(n)$. First, label the edges of T_n by the following. For each node of degree p, with children labeled with $x_0 < x_1 < \ldots < x_{p-1}$, label the edge incident to x_0 by 0, the edge incident to x_1 by 1, and so on. Given these edge labels, The construction is simple: represent each path with the bit string $a_n a_{n-1} \ldots a_0$, where a_j is a bit string of length $\lceil \log j \rceil = \log j + O(1)$, is the binary representation of the label of the edge used in the j'th walk.

Note that the factorial numbers have local properties under swap. In order to apply a swap on this encoding one needs to alter only $O(\log n)$ bits. Suppose that the permutations $\sigma = \sigma(1), \sigma(2), \ldots, \sigma(k), \sigma(k+1), \ldots, \sigma(n)$ and $\pi = \sigma(1), \sigma(2), \ldots, \sigma(k+1), \sigma(k), \ldots, \sigma(n)$ are represented by the binary encoding $X = a_1, a_2, \ldots a_k, a_{k+1} \ldots, a_n$ and $Y = b_1, b_2, \ldots b_k, b_{k+1} \ldots, b_n$, respectively. Clearly, π can be obtained from σ by swapping the element k and $k+1$. Notice that $a_1 = b_1, a_2 = b_2, \ldots, a_{k-1} = b_{k-1}$. This is because the corresponding path representations of the two permutations on T_n walk through the same node at the $k-1$'th walk. Also $a_{k+2} = b_{k+2}, \ldots, a_n = b_n$. This is because the subtrees behind the $k+2$'th layer nodes in the two paths are two copies of the same tree, since their nodes consist of same index sets. Therefore, X and Y differ only at a_k, a_{k+1} and b_k, b_{k+1} substrings. As a consequence of these observations, the bit-string codes for two permutations that differ in adjacent labels only, are different in $O(\log n)$ bits.

6.9 Detailed Proofs for Section 6.5

Here we provided detailed proofs of how to recover BQP from QBall on arbitrary initial states, as well as a partial classification of those states for which QBall obtains the power of BQP.

6.9.1 Review of Exchange Interactions

We show how to use arbitrary initial states to obtain a programmable BQP universal model. This is done by demonstrating a reduction from the exchange interaction model of quantum computation which is already known to be BQP universal.

Here, we first briefly review the exchange interaction model [166, 41, 167], and then describe how to do a reduction from the computation in this model to the ball permuting model of computing on arbitrary initial states. Next, we sketch the proof of universality for the exchange interaction model, which in turn results in BQP universality of ball permuting model on arbitrary initial states.

Consider the Hilbert space $(\mathbb{C}^2)^{\otimes n} =: \mathbb{C}\{0,1\}^n$, with binary strings of length n, $\mathcal{X}_n := \{|x_1\rangle \otimes |x_2\rangle \otimes \ldots \otimes |x_n\rangle : x_j \in \{0,1\}\}$, as the orthonormal computational basis. Exchange interactions correspond to the unitary gates $T(\theta, i, j) = \exp(i\theta E_{(i,j)}) = \cos\theta I + i \sin\theta E_{(i,j)}$, where the operator, $E_{(i,j)}$, called the exchange operator, acts as:

$$E = \frac{1}{2}(I + \sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z)$$

on the i, j slots of the tensor product, and acts as identity on the other parts. More specifically, E is the map:

$$\begin{aligned} |00\rangle &\rightarrow |00\rangle, \\ |01\rangle &\rightarrow |10\rangle, \\ |10\rangle &\rightarrow |01\rangle, \\ |11\rangle &\rightarrow |11\rangle. \end{aligned}$$

The action of E_{ij} is very similar to the permuting operator $L_{(i,j)}$, except that E operates on bits rather than the arbitrary labels of $[n]$. These operators are also known as the Heisenberg couplings, related to the Heisenberg Hamiltonian for spin-spin interactions.

6.9.2 Reduction from Exchange Interactions

Define $\mathcal{X}_n^k := \{|x\rangle : x \in \{0,1\}^n, |x|_H = k\}$ to be the subset of $\mathcal{X}_n$, containing strings of Hamming distance $k \leq n$. Here, $|.|_H$ is the Hamming distance. Also, let $\mathbb{C}\mathcal{X}_n^k$ be the corresponding Hilbert space spanned by these basis.

Theorem 6.9.1. *Given a description of $U = T(\theta_m, i_m, j_m)\ldots T(\theta_2, i_2, j_2)T(\theta_1, i_1, j_1)$, and an initial state $|\psi\rangle \in \mathbb{C}\mathcal{X}_n^k$, there exists an initial $|\psi'\rangle \in \mathbb{C}S_n$, and a ball permuting circuit, with X operators, that can sample from the output of $U|\psi'\rangle$, exactly.*

Proof. We show how to encode any state of $\mathbb{C}\mathcal{X}_n^k$ with states of $\mathbb{C}S_n$. Let $S_{k,n-k}$ be the subgroup of S_n according to the cycles $\{1, 2, \ldots, k\}$ and $\{k+1, k+2, \ldots, n\}$, and denote $|\phi_0\rangle = \frac{1}{\sqrt{k!(n-k)!}} \sum_{\sigma \in S_{k,n-k}} R(\sigma)|123\ldots n\rangle$ be an encoding of the state $|1^k 0^{n-k}\rangle$. Here, 1^k means 1's repeated for k times. This is indeed a quantum state that is symmetric on each the labels of $\{1, 2, \ldots, k\}$ and $\{k+1, k+2, \ldots, n\}$, separately. Any string of Hamming distance k can be obtained by permuting the string $0^k 1^{n-k}$. For any such string x let π_x be such a permutation, and encode $|x\rangle$ with $|\phi(x)\rangle := L_{\pi_x}|\phi_0\rangle$. Therefore, given any initial state $|\psi\rangle := \sum_{x \in \mathcal{X}_n^k} \alpha_x |x\rangle$, pick an initial state $|\psi'\rangle := \sum_{x \in \mathcal{X}_n^k} \alpha_x |\phi(x)\rangle$ in $\mathbb{C}S_n$. Now, given any unitary $U = T(\theta_m, i_m, j_m)\ldots T(\theta_2, i_2, j_2)T(\theta_1, i_1, j_1)$ with T operators, pick a corresponding ball permuting circuit $U' = X(\theta_m, i_m, j_m)\ldots X(\theta_2, i_2, j_2)X(\theta_1, i_1, j_1)$. It can be confirmed that for any $i < j \in [n]$ if $E_{(i,j)}|x\rangle = |x'\rangle$, then $E_{(i,j)}|\phi(x)\rangle = |\phi(x')\rangle$. From this, if $U|\psi\rangle = \sum_{x \in \mathcal{X}_n^k} \beta_x |x\rangle$, then $U'|\psi'\rangle = \sum_{x \in \mathcal{X}_n^k} \beta_x |\phi(x)\rangle$.

It remains to show that given access to the output of $U'|\psi'\rangle$, one can efficiently sample from $U|\psi\rangle$. Suppose that $U'|\psi'\rangle$ is measured in the end, and one obtains the permutation $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$. Then, by outputting a string x by replacing all the labels of

$\{1, 2, \ldots, k\}$ in σ with ones and the other labels with zeros the reduction is complete. The probability of obtaining any string x with this protocol is exactly equal to $|\langle x|U|\psi\rangle|^2$. □

Universal quantum computing is possible by encoding a qubit using three spin 1/2 particles. Suppose that the following initial states are given in $\mathbb{C}\mathcal{X}_3^1$:

$$|0_L\rangle := \frac{|010\rangle - |100\rangle}{\sqrt{2}}$$

and,

$$|1_L\rangle := \frac{2|001\rangle - |010\rangle - |100\rangle}{\sqrt{6}},$$

as some logical encoding of a qubit using three quantum digits. We claim that there is a way to distinguish $|0_L\rangle$ from $|1_L\rangle$ with perfect soundness. These mark the multiplicity space of the space with half Z direction angular momentum and half total angular momentum. First, we should find a way to distinguish between these two states using measurement in the computational basis. Suppose that we have access to k copies of an unknown quantum state, and we have the promise that it is either $|0_L\rangle$ or $|1_L\rangle$, and we want to see which one is the case. The idea is to simply measure the third bit of each copy, and announce it to be 0_L if the results of the k measurements are all 0 bits. If the state has been $|0_L\rangle$, the probability of error in this decision is zero, because $|0_L\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}} \otimes |0\rangle$. Otherwise, we will make a wrong decision with probability at most $(1/3)^k$, which is exponentially small. This is because the probability of reading a 0 in the third bit of $|1_L\rangle$ is 1/3.

Theorem 6.9.2. *There is a way of acting as encoded $SU(2)$ on the span of $\{|0_L\rangle, |1_L\rangle\}$, and also $SU(4)$ on the concatenation of two encoded qubits.*

Proof. (Sketch) according to the analysis of [103, 166], one can look at the Lie algebra of the exchange operators to find encoded $\mathfrak{su}(2)$ algebra on the encoded qubit. Also, we need to take enough commutations such that the action of the designed operators annihilates the two one dimensional spaces spanned by $|000\rangle$ and $|111\rangle$. The authors of [166] prove that there is a way to act as $SU(V(s,m))$ on each invariant subspace $V(s,m)$. Here $V(s,m)$ is the subspace corresponding to total angular momentum s and z direction angular momentum m. Moreover, they prove that the action on two subspaces $V(s_1, m_1)$ and $V(s_2, m_2)$ can be decoupled, unless $s_1 = s_2$, and $m_2 = -m_1$, where the two subspaces are isomorphic. It is almost enough to prove that the state $|0_L\rangle \otimes |0_L\rangle$ is contained in non-isomorphic invariant subspaces. However, this is also true, since $|0_L\rangle \otimes |0_L\rangle$ is completely contained in subspaces with $m = 2$. □

See [52, 168, 244] for similar models with encoded universality. Therefore, this is a nonconstructive proof for the existence of an encoded entangling quantum gate; CNOT for example. Indeed, the actual construction of a CNOT is given in [103] . Notice that for a decision problem, one can formulate quantum computation in such a way that only one qubit needs to be measured in the end, and this can be done by distinguishing $|0_L\rangle$ and $|1_L\rangle$ using measurement in the computational basis. The probability of success in distinguishing between the two bits can also be amplified by just repeating the computation for polynomial number of times, and taking the majority of votes. Also, taking the majority of votes can be done with encoded CNOTs and single qubits gates on a larger circuit, and without loss of generality we can assume that one single measurement on one single qubit is sufficient.

6.9.3 Partial Classification of Quantum Computation on Different Initial States

In the following, it is proved that the ball permuting gates act densely on invariant subspaces corresponding to Young tableaus with two rows or two columns. The proof is based on the bridge lemma and decoupling lemma of reference [28]. As we discuss, conditioned on the existence of a bridge operator, and decoupled dense action on two orthogonal subspace of different dimensionality, the bridge lemma glues the two subspaces into a larger subspace with dense action on it. Also, the decoupling lemma decouples action on two orthogonal subspaces of different dimensionality, given dense action on each of them. Consulting with [175], it is conceivable that these two lemmas have natural generalizations to more than two subspaces and subspaces that have equal dimensionality. We conjecture that using these tools one can prove that the action of ball permuting gates is dense on all invariant subspaces of the symmetric group, even for those which correspond to Young diagrams of more than two rows/columns. We leave this investigation to further work.

In this section, the Lie algebra and the unitary Lie group generated by X operators are used interchangeably. The Hilbert space $\mathbb{C}S_n$ has the decomposition

$$\mathbb{C}S_n \cong \bigoplus_{\lambda \vdash n} V_\lambda \otimes X_\lambda$$

The explicit description of the basis of V_λ as superpositions of the permutation basis can be found in [158].

Let G be the unitary group generated by these $X(\theta, k) = \exp(i\theta L_{(k,k+1)})$ operators. As described earlier, the space tangent to the identity element of G is a Lie algebra, g, which contains $L_{(k,k+1)}$ for all $k \in [n-1]$, and is close under linear combination over $\mathbb{R}$, and the Lie commutator $i[\cdot,\cdot]$. The objective is to show that for any $\lambda \vdash n$ with two rows or two columns, and any element U of $SU(V_\lambda)$, there is an element of G that is arbitrarily close to U.

The proof is presented inductively. First of all, for any n, the irreps V_n and $V_{1,1,1,\ldots,n}$ are one dimensional, and the action of $x \in G$ is to add an overall phase. However, observing the structure of YY basis for these irreps, the action of G on the joint blocks $V_n \oplus V_{(1,1,1,\ldots,1)}$ cannot be decoupled, and the projection of G onto these subspaces is diagonal, and moreover isomorphic to the group $e^{i\theta} \times e^{-i\theta} : \theta \in \mathbb{R}$. Intuitively, these are Bosonic and Fermionic subspaces, where an exchange $L_{(k,k+1)}$ of particles results in a $+1$ and -1 overall phase, respectively.

For $n = 2$, the only invariant subspaces are V_2 and $V_{(1,1)}$, and we know the structure of these irreps from the last paragraph

$$\mathbb{C}S_2 \cong V_2 \oplus V_{(1,1)}, \qquad G \twoheadrightarrow e^{i\theta} \times e^{-i\theta} : \theta \in \mathbb{R}.$$

For $n = 3$, the decomposition is according to

$$\mathbb{C}S_3 \cong V_3 \oplus V_{(1,1,1)} \oplus V_{(2,1)} \otimes X(2).$$

Here, $X(2)$ is a two dimensional multiplicity space. There are two standard $(2,1)$ tableaus and therefore $V_{(2,1)}$ is also two dimensional. Observing the YY basis the two generators $L_{(1,2)}$ and $L_{(2,3)}$ take the matrix forms

$$L_{(1,2)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and,

$$L_{(2,3)} = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}.$$

The basis of the matrix are marked with the two standard Young tableaus of shape $(2,1)$. The first basis corresponds to the numbering $(1,2;3)$ and the second one corresponds to $(1,3;2)$. Here, the rows are separated by semicolons. The following elements of the Lie algebra g generate $\mathfrak{su}(V_{(2,1)})$ and annihilate the two Bosonic and Fermionic subspaces

$$\frac{1}{2\sqrt{3}}[L_{(1,2)}, [L_{(1,2)}, L_{(2,3)}]] = 0 \oplus 0 \oplus \sigma_x \otimes I,$$

$$\frac{i}{\sqrt{3}}[L_{(1,2)}, L_{(2,3)}] = 0 \oplus 0 \oplus \sigma_y \otimes I,$$

and,

$$\frac{1}{6}[[L_{(1,2)}, [L_{(1,2)}, L_{(2,3)}]], [L_{(1,2)}, L_{(2,3)}]] = 0 \oplus 0 \oplus \sigma_z \otimes I.$$

This implies the denseness of G in $1 \times 1 \times SU(V_{(2,1)})$. Therefore, we obtain a qubit coupled to the multiplicity space, placed in a superposition of the one dimensional Bosonic and Fermionic subspaces. So, projecting onto a subspace like $V_{(2,1)} \otimes |\psi\rangle$, for $|\psi\rangle \in X(2)$, we obtain a qubit.

We use this result as the seed of an induction. The upshot is to add boxes to $(2,1)$ one by one, in a way that the partitions remain with two rows or two columns. At each step, we use the branching rule to combine the blocks together to larger and larger special unitary groups. In the course of this process, we use two important tools, called the bridge lemma, and decoupling lemma:

Lemma 6.9.1. *(Aharonov-Arad [28]) let A and B be two orthogonal subspaces, with non-equal dimensions,* $\dim A < \dim B$*:*

- *(Bridge) if there is some state $|\psi\rangle \in A$, and a (bridge) operator $V \in SU(A \oplus B)$, such that the projection of $V|\psi\rangle$ on B is nonzero, then the combination of $SU(A)$, $SU(B)$, and V is dense in $SU(A \oplus B)$.*
- *(Decoupling) suppose for any elements $x \in SU(A)$ and $y \in SU(B)$, there are two corresponding sequences I_x and I_y in G, arbitrarily close to x and y, respectively, then the action of G on $A \oplus B$ is decoupled, i.e., $SU(A) \times SU(B) \subseteq G$.*

See [31, 29] for more similar results. Intuitively, what bridge lemma says is that given two subspaces, with one of them larger than the other, dense action each, along with a bridge between them, implies denseness on the combined subspace. That is a bridge glues them to a larger special group. The condition of different dimensions is a crucial requirement for the application of this lemma. The decoupling lemma, on the other hand, states that given dense action on two subspaces, as long as they have different dimensionality, there is way of

acting on the two subspaces independently. Again, in this case non-equal dimensionality is important. For example, suppose that $\dim A = \dim B$, then the action $x \times \bar{x} : x \in SU(A)$, cannot be decoupled. Here $\bar{x}$ is the complex conjugate of x, i.e., entries $\bar{x}$ as a matrix are complex conjugates of corresponding entries of matrix x. In order to see this, just notice that after finite compositions, the general form of elements generated in this way is $(x_1 x_2 \dots x_n) \times \overline{(x_1 x_2 \dots x_n)}$, and an identity action on the left part forces identity action on the right part of the Cartesian product.

Next, we show that the lemma along with the branching rule, force denseness on all irreps corresponding to partitions of two rows or two columns. We will take care of the case with two rows. The situation with two columns is similar. As a way of induction, suppose that, for any $m < n$, for any $\lambda = (\lambda_1 \geq \lambda_2) \vdash m$, the projection of G on λ is dense in $SU(V_\lambda)$. The objective is to prove denseness for any partition $\mu \vdash n$.

This is true for $(2,1)$, as showed above. For the sake of illustration, we prove this for $n = 4$. The partitions (4) is immediate, because this is one dimensional. Also, the partition $(2,2)$ is immediate, since the branching rule, under the action of S_3 is

$$V_{(2,2)} \cong V_{(2,1)},$$

That is the only removable box from $(2,2)$ is the last box, and in the YY basis for $(2,2)$, this last box can contain the symbol 4 only. So, the same operators of S_3 act densely on this subspace.

The situation with the partition $(3,1)$ is a little different. Analyzing the hook lengths, $V_{(3,1)}$ has dimension 3, and the branching rule involves the direct sum of partitions $(2,1)$ and (3)

$$V_{(3,1)} \cong V_{(2,1)} \oplus V_{(3)}.$$

Where, $V_{(2,1)}$ is two dimensional, and $V_{(3)}$ is one dimensional, and therefore, they have non-equal dimensions, and also their direct sum adds up to dimension 3. From, the analysis of S_3 we know that independent $SU(2)$, and $SU(1) = \{1\}$ is possible on these irreps. It suffices to find a bridge operator in $SU(V_{(2,1)} \oplus V_{(3)})$. In the first glance, the operator $L_{3,4} \in g$ sounds like a suitable choice. However, there is a problem with this: the restriction of $L_{3,4}$ on $V_{(3,1)}$ is not traceless, and therefore the image under exponentiation does not have unit determinant. Therefore, a wise choice for a bridge operator is $i[L_{(2,3)}, L_{(3,4)}]$. Looking at the actual matrices, restricted to the YY basis of $(3,1)$, one finds $i[L_{(2,3)}, L_{(3,4)}]$, as a suitable bridge, that is nice and traceless

$$i \begin{pmatrix} 0 & \sqrt{2} & -\sqrt{\frac{2}{3}} \\ -\sqrt{2} & 0 & \sqrt{\frac{1}{3}} \\ \sqrt{\frac{2}{3}} & -\sqrt{\frac{1}{3}} & 0 \end{pmatrix}.$$

Here the matrix is written in the basis corresponding to the tableaus $(1,2,3;4), (1,2,4;3)$ and $(1,3,4;2)$. The bridging is between the $(1,2)$ and $(2,1)$ elements of the matrix. Thereby, the bridge lemma implies the desired denseness.

For general n, two situations can happen, either the partition under analysis is of the form $(\nu, \nu) = (n/2, n/2)$ (for even n of course), or not. In the first case, the situation is

similar to the partition $(2,2)$ of $n=4$. Thereby, restricted to S_{n-1}

$$V_{(\nu,\nu)} \cong V_{(\nu,\nu-1)},$$

and based on the induction hypothesis the image of G is already dense in the subspace. In the second case, also two cases can happen: either the partition has the form $\mu=(\nu+1,\nu)$, with $2\nu+1=n$, or not. In the first case, the branching rule is according to

$$V_{(\nu+1,\nu)} \cong V_{(\nu,\nu)} \oplus V_{(\nu+1,\nu-1)}.$$

The space $V_{(\nu,\nu)}$ corresponds to all YY basis corresponding to tableaus, wherein the index n is located in the last box of the first row. Therefore, the index $n-1$ in all of the tableaus of (ν,ν) is located in the last box of the second column, because this is the only removable box available. For simplicity, let's call this space V_1. The YY bases of $V_{(\nu+1,n-1)}$ correspond to all the tableaus of $(\nu+1,\nu)$, where the index n is located in the last box of the second row. In this space, the location of the index $n-1$ is either in the last box in the first row or in the box right at the left of the last box in the second row. A coarser stratification of the states in $V_{(\nu+1,\nu-1)}$ is by grouping the YY basis according to the location of $n-1$. Let V_2 be the first one, and V_3 the second one. Therefore, YY bases of $V_{(\nu+1,\nu)}$ can be grouped in three ways, V_1, V_2, V_3, corresponding to all the ways that one can remove two boxes from the original $V_{(\nu+1,\nu)}$. Again, a neat candidate for a bridge is $L_{(n-1,n)}$. Taking a closer look at the operator $L_{(n-1,n)}$, it can be decomposed according to

$$L_{(n-1,n)} = \sum_{|j\rangle\in V_3} |j\rangle\langle j| + \frac{1}{2} \sum_{\substack{k':k \\ |k\rangle\in V_1 \\ |k'\rangle\in V_2}} |k\rangle\langle k| - |k'\rangle\langle k'| + \sqrt{\frac{3}{2}} \sum_{\substack{k':k \\ |k\rangle\in V_1 \\ |k'\rangle\in V_2}} |k\rangle\langle k'| + |k'\rangle\langle k|$$

$|j\rangle$, $|k\rangle$, and $|k'\rangle$, of V_1, V_2, and V_3 are the corresponding orthonormal basis in the spaces. Notice that the space V_1 is isomorphic to V_2, and $k:k'$, refers to this isomorphism. Clearly, the restriction of $L_{(n-1,n)}$ to this block is not traceless, and indeed $tr_{V_{(\nu+1,\nu)}} = \dim V_3 = \dim V_{(\nu+1,\nu-2)}$.

Now, we use the decoupling lemma of Aharonov-Arad. $V_{(\nu+1,\nu)}$ and $V_{(\nu,\nu)}$ have different dimensionality, and also, due to the induction hypothesis the operators can act as the special unitary group on each of them. Thereby, there is a way to act as $x \oplus 0$ on the joint space $V_{(\nu,\nu)} \oplus V_{(\nu+1,\nu-1)}$, for some traceless element $x \in \mathfrak{su}(V_{(\nu,\nu)})$. Therefore, $x|j\rangle = 0$ and $x|k'\rangle$, for all $|j\rangle \in V_3$, $|k'\rangle \in V_2$. And denote $|xk\rangle := x|k\rangle$, for $|k\rangle \in V_1$. Taking the commutator $i[x, L_{(n-1,n)}]$

$$i[x, L_{(n-1,n)}] = \frac{i}{2} \sum_{\substack{k':k \\ |k\rangle\in V_1 \\ |k'\rangle\in V_2}} |xk\rangle\langle k| - |k\rangle\langle xk| + i\sqrt{\frac{3}{2}} \sum_{\substack{k':k \\ |k\rangle\in V_1 \\ |k'\rangle\in V_2}} |xk\rangle\langle k'| - |k'\rangle\langle xk|.$$

Clearly, this operator is traceless, Hermitian, and also one can choose x in such a way that the bridging term in the second sum is nonzero.

Given the above proof for the case $V_{(\nu+1,\nu)}$, we will use a similar technique to take care of the situation $V_{(p,q)}$, where $p > q+1$, and $p+q=n$. Again, the branching rule is:

$$V_{(p,q)} = V_{(p,q-1)} \oplus V_{(p-1,q)}.$$

The space $V_{(p,q-1)}$ corresponds to all YY bases that correspond to the tableaus where the index n is located at the last box of the first row. In this space, the index $n-1$ is either located at the left side of the box containing n, or it is located in the last box of the second row. Call the space corresponding to the first (second) one V_1 (V_3). $V_{(p-1,q)}$ corresponds to all YY bases of tableaus with index n is located at the last box of the second row. In this space, the index $n-1$ is either located at the left side of the box containing n, or it is located in the last box of the first row. Call the first space V_2 and the second one V_4. Again, write the decomposition of $L_{(n-1,n)}$, accordingly

$$L_{(n-1,n)} = \sum_{|j\rangle \in V_1} |j\rangle\langle j| + \sum_{|j\rangle \in V_2} |j\rangle\langle j| + \alpha(p,q) \sum_{\substack{k':k \\ |k\rangle \in V_3 \\ |k'\rangle \in V_4}} |k\rangle\langle k| - |k'\rangle\langle k'|$$
$$+ \beta(p,q) \sum_{\substack{k':k \\ |k\rangle \in V_3 \\ |k'\rangle \in V_4}} |k\rangle\langle k'| + |k'\rangle\langle k|$$

Here

$$\alpha(p,q) = \frac{1}{p-q+1}$$

and,

$$\beta(p,q) = \sqrt{1 - \frac{1}{(p-q+1)^2}}.$$

Once again, V_2 is isomorphic to V_3, and $k : k'$ denotes the correspondence between elements of the two spaces. Once again, we use the decoupling lemma, which asserts the existence of elements like $X := x \oplus 0$, and $Y := 0 \oplus y$, on $V_{(p,q-1)} \oplus V_{(p-1,q)}$, for every $x \in \mathfrak{su}(V_{(p,q-1)})$ and $y \in \mathfrak{su}(V_{(p,q-1)})$. A bridge between V_3 and V_4 is needed, in such a way that the bridge annihilates both V_1 and V_2. A candidate for a bridge is $[Y,[X,L_{(n-1,n)}]]$. However, it can be easily shown that the element $i[X,L_{(n-1,n)}]$ will also work. The operator X annihilates everything in V_2 and V_4. Therefore, taking the commutator, the second sum is annihilated, and also, all the remaining terms are traceless and one can find x in such a way that the bridge part is nonzero. All the above results also apply to the tableaus with two columns.

6.10 Open Problems and Further Directions

We leave open a number of interesting problems:

1. A natural open problem is to finish classifying the power of QBall on all input states. While we nearly completed this task, the tools we used for the classification are restricted in the sense that they can only take care of special subspaces and not the

others. Another natural open problem is to determine if QBall with the starting state $|12\ldots n\rangle$ is contained in DQC1 as a decision class.

2. It remains open to generalize the result of Section 6.4 to arbitrary quantum models with Hilbert spaces based on group algebras. As discussed, such a generalization relies on the ability to do the following: given the description of the generators of a group, encode the elements of an arbitrary element with binary strings with nearly $\log |G|$ bits so that the action of group elements on each other is implementable by reversible circuits that affect only $O(\log \log |G|)$ bits at a time. We established this property for the symmetric group using the factorial number system. Are there other groups with this property?

3. In Section 6.4 we show that if the quantum ball permuting model has access to arbitrary initial states, then there is a way to efficiently sample from standard quantum circuits. The construction is based on encoding of qubits using superpositions over permutation states. More precisely n qubits can be encoded using a superposition over permutations of $3n$ labels. However, a drawback of this construction is that it is not composable, in the sense that the encoding of k qubits is not obtained by taking a k-fold tensor product of the encoding of a single qubit. Instead, to encode k qubits, we use $3k$ qubits and symmetrize over all the labels representing 0's and all the labels representing 1's. A natural question is to give a composable encoding of qubits in this model, or prove such an encoding is impossible.

4. In section 6.6 we introduced the complexity class SampTQP, in which we start with the maximally entangled initial state $\sum_x |x\rangle|x\rangle$, we apply an arbitrary quantum circuit to the left half of this state and at the end we measure both halves. We show that the power of this model is intermediate between DQC_1 and BQP. It is open to further classify the power of this class.

5. Is there an encoding of the Young-Yamanouchi basis of arbitrary tableaux into bit strings which is both extremely compressed (i.e. using nearly the information theoretically optimal number of bits) and local (i.e. exchanging two labels is an operation which can be performed on $O(\log n)$ bits). We believe that this should be achievable. Indeed given a circuit C, we can compute the average trace of C over the partitions $\sum_\lambda p_\lambda \operatorname{Tr}_\lambda(C)$ as this is simply equal to $\langle 123\ldots n|C|123\ldots n\rangle$. So it would be strange (but not logically impossible) if one could not compute the individual $\operatorname{Tr}_\lambda(C)$'s.

6. In section 6.5.2 we partially classify the computational power of the ball permuting model on arbitrary initial states. We prove that the unitary group generated by this model is as large as possible if the model starts from the initial states corresponding to Young diagrams with two rows or two columns. We conjecture that this is true for arbitrary irreducible representations. The main difficulty in proving this is that the bridge lemma (Lemma 6.9.1) works only if the subspaces are of *different dimensions.* However, there are cases where two subspaces of equal dimensionality take part in a single branching rule, so the bridge lemma is not applicable. An additional difficulty is that the action of ball permuting gates on different irreducible subspaces can be coupled. For example, if λ^T is the transpose of the partition λ then if a sequence of ball permuting gates apply the unitary U on $V(\lambda)$ then the same sequence applies the

unitary U^* to $V(\lambda^T)$ after some change of basis. However, we conjecture this is the only coupling possible in our decomposition. Furthermore, even with such coupling between irreps the model we obtain can be BQP universal. For example, consider the situation where the initial state is $|\psi\rangle = \frac{1}{\sqrt{2}}(|x\rangle+|x'\rangle)$ where $|x\rangle$ and $|x'\rangle$ are in separate irreps. Then the coupled action of the form $U \oplus U^\star$ maps $|\psi\rangle$ to $\frac{1}{\sqrt{2}}(U|x\rangle + U^\star|x'\rangle)$, so $\langle\psi|C|\psi\rangle = \frac{1}{2}(\langle x|U|x\rangle + \langle x'|U^\star|x'\rangle) = \Re\langle x|U|x\rangle$. Note the problem of reading a real entry of a quantum circuit is already known to be BQP-complete, so approximating $\langle\psi|C|\psi\rangle$ is BQP-complete. We leave the full classification to future work.

Part II

The space above BQP

Chapter 7

The space "just above" BQP

In this chapter, we explore the space "just above" BQP by defining a complexity class naCQP (non-adaptive Collapse-free Quantum Polynomial time) which is larger than BQP but does not contain NP relative to an oracle. The class is defined by imagining that quantum computers can perform (non-adaptive) measurements that do not collapse the wavefunction. This non-physical model of computation can efficiently solve problems such in SZK (such as Approximate Shortest Vector) which are believed to be intractable for quantum computers. Furthermore, it can search an unstructured N-element list in $\tilde{O}(N^{1/3})$ time, but no faster than $\Omega(N^{1/4})$, and hence cannot solve NP-hard problems in a black box manner. In short, this model of computation is more powerful than standard quantum computation, but only slightly so. This is surprising as most modifications of BQP increase the power of quantum computation to NP or beyond.

This chapter is based on joint work with Scott Aaronson, Joseph Fitzsimons, and Mitchell Lee [13].

7.1 Introduction

Quantum computers are believed to be strictly more powerful than classical computers, but not so much more powerful that they can solve NP-hard problems efficiently. In particular, as discussed in Section 2.1.3, it is known that BQP does not contain NP "relative to an oracle." This means that there is some "black box" problem $\mathcal{O}$ for which $\mathsf{BQP}^{\mathcal{O}} \not\supset \mathsf{NP}^{\mathcal{O}}$. (For more information about the terminology, see [38] as well as Section 2.1.1.) On the other hand, as discussed in section 2.4.1 many seemingly innocuous modifications of quantum mechanics—for example, allowing nonlinear transformations, non-unitary transformations, postselection, or measurement statistics based on the pth power of the amplitudes for $p \neq 2$—increase the power of quantum computation drastically enough that they can solve NP-hard problems (and even #P-hard problems) efficiently [22][6]. As a result, it is difficult to find natural complexity classes which are bigger than BQP but which don't contain NP. Quantum mechanics appears to be an "island in theoryspace" in terms of its complexity-theoretic properties [3].

In this chapter, we explore a natural modification of quantum mechanics to obtain a complexity class which is only "slightly more powerful" than BQP. In quantum mechanics, when a system is measured, the state of the system "collapses" to its observed value; one cannot observe a quantum system without perturbing it. Here we consider the power of quantum computers which can also make "non-collapsing measurements," which are identical

to usual quantum measurements except that they do not perturb the state. We call the class of problems decidable in polynomial time in this model CQP, which stands for "Collapse-free Quantum Polynomial time." We additionally consider a weaker version of this model, naCQP (non-adaptive CQP) in which the quantum operations performed must be independent of the non-collapsing measurement outcomes.

We show that quantum computers equipped with this power (even in the non-adaptive case) can solve any problem in SZK in polynomial time in a black-box manner[1]. Since standard quantum computers cannot solve SZK-hard problems in a black-box manner [2], this implies that there is an oracle $\mathcal{O}$ for which $\mathsf{BQP}^{\mathcal{O}} \neq \mathsf{naCQP}^{\mathcal{O}}$. This is evidence that quantum computation with non-collapsing measurements is more powerful than standard quantum computation. Furthermore, we upper bound the power of both CQP and naCQP by showing that $\mathsf{naCQP} \subseteq \mathsf{CQP} \subseteq \mathsf{BPP}^{\mathsf{PP}}$, so both naCQP and CQP are in the counting hierarchy. In comparison the best known classical upper bound for BQP is AWPP which is contained in PP [119][24].

We also demonstrate that if (even non-adaptive) non-collapsing measurements are possible, then there is a quantum algorithm that searches an unstructured list of N elements in $\tilde{O}(N^{1/3})$ time. Furthermore any such algorithm takes at least $\Omega(N^{1/4})$ time in the non-adaptive case. While the upper bound is simple, the proof of the lower bound uses a hybrid argument [56] and properties of Markov chains. We conclude that naCQP does not contain NP relative to an oracle. To our knowledge this represents the only known complexity class larger than BQP which provably does not admit polynomial time black-box algorithms for NP-hard problems. This is what we mean when we say naCQP is only "slightly more powerful" than BQP. Proving the analogous lower bound for CQP remains open, so it is possible that CQP could be more powerful than naCQP.

Note that introducing non-collapsing measurements into quantum mechanics allows for many strange phenomena. In particular, it allows for faster-than-light communication, it allows for (approximate) quantum cloning[2], and it renders quantum query complexity and quantum communication complexity meaningless. We describe these strange consequences of non-collapsing measurements in detail in Section 7.8. For this reason, we are not suggesting that "non-collapsing measurements" should be considered seriously as an amendment to quantum theory. Rather we are simply showing that non-collapsing measurements have interesting complexity-theoretic properties - namely, that they can be used to define an complexity class which is "just above" BQP.

7.2 Relation to Prior Work

Our work is inspired by previous work on quantum computing with hidden variables by Aaronson [5]. Aaronson defines a class DQP ("Dynamical Quantum Polynomial Time") by imagining a hidden variable theory is true, and that an experimenter can view the evolution of the hidden variables in real time. Additionally, he requires that the quantum operations are non-adaptive to the hidden variable values (similar to our class naCQP). He shows that with this power one can search in $\tilde{O}(N^{1/3})$ time and solve any problem in SZK in

[1] In particular this power would allow one to break lattice-based cryptography, which so far has proven immune to quantum attack [107].

[2] Note, however, that this only arises when the quantum operations can depend on the non-collapsing measurement results. Our definition of naCQP does not allow cloning due to the non-adaptivity restriction. In contrast the class CQP does admit cloning, so might be a more powerful computational model. We discuss this issue in detail in Section 7.8, and describe a related open problem in Section 7.15.

Table 7.1: Comparison between BQP, naCQP, CQP and DQP

Property	BQP	naCQP	CQP	DQP
Contains SZK	Unknown	Yes	Yes	Yes
Contains SZK^O $\forall O$	No	Yes	Yes	Yes
Upper Bound for Search	$O(N^{1/2})$	$\tilde{O}(N^{1/3})$	$\tilde{O}(N^{1/3})$	$\tilde{O}(N^{1/3})$
Known Lower Bound for Search	$\Omega(N^{1/2})$	$\Omega(N^{1/4})$	$\Omega(1)$	$\Omega(1)$
Upper Bound	AWPP	$\mathsf{BPP}^{\mathsf{PP}}$	$\mathsf{BPP}^{\mathsf{PP}}$	EXP

polynomial time. He additionally claims one cannot search in faster than $\Omega(N^{1/3})$ time in this model. Unfortunately, there is an error which invalidates his proof of the lower bound for search. For the interested reader, we describe this error in Section 7.12 and correct it for a modified version of the computational model in Section 7.13. Proving the lower bound for search under Aaronson's original computational model is challenging because we have few examples of working hidden variable theories, and therefore have little understanding of how hidden variable values could correlate over time. Note, however, that an $\Omega(N^{1/3})$ lower bound for search might hold even for Aaronson's original model.

The classes CQP and naCQP, which we define by imagining one can perform (non-adaptive) non-collapsing measurements, seem incomparable to DQP - we do not know if either $\mathsf{CQP} \subseteq \mathsf{DQP}$, nor if $\mathsf{DQP} \subseteq \mathsf{CQP}$. However, we suspect that naCQP is a weaker class than DQP for several reasons. First, we can prove a polynomial lower bound for search in naCQP, which we don't know how to do in DQP. Second, we can prove an upper bound that $\mathsf{naCQP} \subseteq \mathsf{BPP}^{\mathsf{PP}} \subseteq \mathsf{PSPACE}$. In contrast the best known upper bound for DQP is EXP. Table 7.1 summarizes the relationship between BQP, naCQP, CQP and DQP.

7.3 Definition of CQP and naCQP

We assume the reader is familiar with the standard definition of BQP and the basics of quantum computing; for an introduction to this topic see [199] and Section 2.1. We now give a formal definition of our model of quantum computing with non-collapsing measurements.

Let $\mathcal{Q}$ be an oracle that takes as input a quantum circuit $C = (U_1, M_1, \cdots, U_T, M_T)$ and an integer $\ell \geq 0$. Here each U_i is a unitary operator on ℓ qubits composed of gates from some finite universal gate set $\mathcal{U}$, and each M_i is a standard (collapsing) measurement of zero or more qubits in the computational basis. Define a (random) sequence $\{|\psi_t\rangle\}_{t=0}^T$ of quantum states by $|\psi_0\rangle = |0\rangle^{\otimes \ell}$ and for $t > 0$, $|\psi_t\rangle$ is the resulting (random) pure state obtained when measurement M_t is applied to $U_t|\psi_{t-1}\rangle$. Note that we imagine the state of the system $|\psi_t\rangle$ is a (random) pure state for $0 \leq t \leq T$. The oracle $\mathcal{Q}$ samples the sequence $\{|\psi_t\rangle\}_{t=0}^T$ (note that the random variables $|\psi_t\rangle$ are not independent), measures $|\psi_t\rangle$ in the computational basis for every t independently, and outputs the $T+1$ measurement results, which we label $v_0, v_1, \ldots v_T$, respectively. The output of $\mathcal{Q}$ is an element of $(\{0,1\}^\ell)^{T+1}$. Note that once the $|\psi_t\rangle$ are fixed, the $T+1$ measurement results are independent, however since the $|\psi_t\rangle$ are correlated, the measurement outcomes may be correlated.

naCQP (non-adaptive Collapse-free Quantum Polynomial-time) is then defined as the class of all languages that can be recognized in polynomial time by a deterministic Turing machine with one query to $\mathcal{Q}$, with error probability at most $\frac{1}{3}$. Note that because the base

machine is polynomially bounded, the circuit C with which it queries $\mathcal{Q}$ must be polynomially sized. Furthermore, since the base machine can use the oracle to output coin flips, it makes no difference if we define the base machine to be deterministic or randomized. This class contains BQP, because one can always query the oracle $\mathcal{Q}$ with a BQP circuit, and then ignore all output except the final measurement outcome. The constant $\frac{1}{3}$ is arbitrary: we can decrease the error probability arbitrarily close to 0 by repetition, which can be accomplished by packing multiple copies of a quantum circuit into a single call to $\mathcal{Q}$. Furthermore, it turns out that the definition of naCQP is not affected by the choice of universal gate set $\mathcal{U}$; this is a consequence of the Solovay-Kitaev Theorem. The proof can be found in Section 7.14.

We can think of the $T+1$ measurement samples from $\mathcal{Q}$ as the results of *non-collapsing* measurements on the state vector, which give information about the state without changing it. For instance, let $|\psi_1\rangle = U_1|0\rangle^{\otimes \ell}$, let M_1, M_2 and M_3 be empty measurements, and let U_2, U_3 be the identity. Then the oracle $\mathcal{Q}$ will output the result of three independent non-collapsing measurements of $|\psi_1\rangle$ in the computational basis. The key point is that the oracle's samples do not disturb the state of the system; only the unitary operators U_i and collapsing measurements M_i do. The oracle $\mathcal{Q}$ gives us information about the intermediate stages of the quantum computation without collapsing the state; this is what gives naCQP additional power over BQP.

Note that by requiring the quantum circuit C to be specified up front, we have enforced the condition that the circuit is non-adaptive to the non-collapsing measurement outcomes (hence the name naCQP). To define CQP, we consider the case where the base machine can query the oracle $\mathcal{Q}$ adaptively. That is, the base machine can first specify $U_1, M_1, \ldots U_t, M_t$ and receive samples $v_1 \ldots v_t$, then based on those samples select $U_{t+1}, M_{t+1} \ldots U_{t'}, M_{t'}$ and receive samples $v_{t+1} \ldots v_{t'}$, etc. CQP is then defined analogously to be the class of languages which can be decided in polynomial time with adaptive queries to $\mathcal{Q}$, with error probability at most $1/3$. This class captures the power of generic computations with non-collapsing measurements.

Note that we explicitly allow for intermediate (collapsing) measurements in our model. In the definition of BQP, the principle of deferred measurement tells us that this is not necessary; the power of standard quantum computers is unchanged by the inclusion of intermediate collapsing measurements. However, in our model this makes a crucial difference. Indeed, suppose that we did not allow for intermediate collapsing measurements; then this model would be simulable in BQP with a polynomial amount of overhead. If there are no intermediate measurements M_i, then $|\psi_t\rangle = U_t U_{t-1} \ldots U_1|0\rangle^{\otimes \ell}$ are no longer random variables but are deterministic pure states, each preparable with a polynomially sized quantum circuit. So a BQP machine could simply prepare $|\psi_1\rangle$ and measure it, then prepare $|\psi_2\rangle$ from scratch and measure it, etc. to obtain the samples $v_0, \ldots, v_T$. This would incur at most quadratic overhead.

When we add intermediate measurements into our model, this simulation strategy no longer works. Indeed, suppose that we performed measurement M_1 to obtain a random state $|\psi_1\rangle$. If we wanted to reproduce this state with a BQP machine, we could try applying M_1 to $U_1|0\rangle^{\otimes \ell}$. However, it might be that the probability of obtaining the same outcome for M_1 is exponentially small, and hence the BQP machine could not prepare another copy of $|\psi_1\rangle$ in polynomial time.

In short, the power of this model comes from the fact that we can perform intermediate measurements which collapse the wave function, and afterwards we can examine the resulting pure state $|\psi_t\rangle$ (which might not be efficiently preparable with a BQP machine) using

multiple non-collapsing measurements. In the next section we will show how to leverage these properties to solve any problem in SZK in polynomial time.

7.4 SZK is contained in naCQP

We will now describe how to use the peculiarities of non-collapsing measurements to solve any problem in SZK in polynomial time. The proof uses essentially the ideas of Aaronson [5], with minor simplifications.

SZK was originally defined as the class of languages admitting statistical zero-knowledge proofs. The precise definition of a statistical zero-knowledge proof can be found in [211], but it is not important here. SZK contains important problems such as Graph Isomorphism and Approximate Shortest Vector. It has been a long-standing open problem whether or not these problems can be solved in quantum polynomial time. Ettinger, Høyer and Knill showed that Graph Isomorphism (and indeed any hidden subgroup problem) can be solved in a black box manner with a polynomial number of queries to the black box, but with exponential post-processing time [108]. More recently, Babai gave a quasipolynomial time classical algorithm for graph isomorphism [40]. Note however that even if Graph Isomorphism is in P, it is possible that problems in SZK lie outside of BQP, as Graph Isomorphism is not known to be complete for the class. For example, thus far the Approximate Shortest Vector problem is not known to admit any subexponential quantum algorithms, despite attempts to create such algorithms [107]. On the other hand, Aaronson [2] showed that BQP does not admit a black-box algorithm for the collision problem, and hence there is an oracle relative to which SZK is not in BQP.

In contrast, we show that quantum computers with non-collapsing measurements can solve any problem in SZK efficiently, i.e. SZK $\subseteq$ naCQP. It is enough to prove that Statistical Difference, a problem shown in [211] to be SZK-complete, is in naCQP. The statistical difference problem is to determine, for two functions $P_0, P_1 : \{0,1\}^n \to \{0,1\}^m$ specified by classical circuits, whether the distributions of $P_0(X), P_1(X)$ for uniformly random X are close or far. Here, two distributions are "close" if their total variation distance is less than $\frac{1}{3}$ and they are "far" if their total variation distance is more than $\frac{2}{3}$.

We now show how to solve this efficiently if we have access to non-collapsing measurements.

Theorem 7.4.1. *The Statistical Difference problem can be solved in polynomial time in* naCQP, *and therefore* SZK $\subseteq$ naCQP.

Proof. By the Polarization Lemma of Sahai and Vadhan [211, Lemma 3.3], we can assume that the distributions $P_0(X)$ and $P_1(X)$ have total variation distance less than 2^{-n^c} or more than $1-2^{-n^c}$, for any constant c. For now, assume that the distributions have total variation distance equal to either 1 or 0.

Our algorithm for the statistical difference problem is as follows. Prepare the state

$$\frac{1}{2^{(n+1)/2}} \sum_{b\in\{0,1\}, x\in\{0,1\}^n} |b\rangle|x\rangle|P_b(x)\rangle.$$

Now, measure the third register with a collapsing measurement to obtain a state $|\phi\rangle$ on the first two registers. If the distributions P_0, P_1 have total variation distance 1, then $|\phi\rangle$ will be of the form $|b\rangle|\psi\rangle$ for some b and $|\psi\rangle$. On the other hand, if they have total variation distance

0, then $|\phi\rangle$ will be an equal superposition $\frac{1}{\sqrt{2}}(|0\rangle|\psi_0\rangle + |1\rangle|\psi_1\rangle)$ where $|\psi_0\rangle$ and $|\psi_1\rangle$ have unit norm. We can distinguish the two cases by now repeatedly performing non-collapsing measurements and examining the value of the first register. If P_0, P_1 have total variation distance 1, then all of these measurements will give the same value b; if P_0 and P_1 have total variation distance 0, then each of these measurements will independently give 0 with probability $\frac{1}{2}$ and 1 with probability $\frac{1}{2}$. We can distinguish the two cases with probability 3/4 by performing three non-collapsing measurements and looking at whether or not they yielded identical values of the first register.

Furthermore, the fact that the total variation distances are merely exponentially close to 0 or 1, rather than actually being equal to 0 or 1, makes little difference. One can show that the probability of seeing the same measurement outcome three times is at most $\frac{1}{4} + O(2^{-n^c})$ if P_0 and P_1 are exponentially close and at least $1 - O(2^{-n^c})$ if P_0 and P_1 are exponentially far apart. We provide a detailed proof of this fact in Section 7.9. Therefore our algorithm will have error probability at most 1/3.

□

Hence SZK is in naCQP, and furthermore we can solve SZK problems in naCQP in a black box manner, i.e. relative to any oracle. Since [2] has the result that $\mathsf{SZK} \not\subset \mathsf{BQP}$ relative to an oracle, we have the immediate corollary[3]:

Corollary 7.4.2. *There exists an oracle $\mathcal{O}$ such that* $\mathsf{naCQP}^{\mathcal{O}} \neq \mathsf{BQP}^{\mathcal{O}}$.

7.5 Search in $\tilde{O}(N^{1/3})$ time

Suppose that we are given query access to a function $f : \{0,1\}^n \to \{0,1\}$ such that the preimage $f^{-1}(1)$ contains exactly one element, x. In the classical randomized computational model, we can find x in $O(N)$ time, where $N = 2^n$, but no faster. In the quantum computational model, on the other hand, we can find x in $O(N^{1/2})$ time using Grover's search algorithm [140], but no faster [56].

Here we show that quantum computers equipped with (non-adaptive) non-collapsing measurements can search in $\tilde{O}(N^{1/3})$ time, where the tilde hides factors in $\log N$. The basic idea is to run $N^{1/3}$ Grover iterations, and then make $N^{1/3}$ non-collapsing measurements of the resulting state. Then with high probability the the marked item will be seen. This is a simplification of the proof given in [5, Theorem 10] for DQP. We now formalize this idea below:

Theorem 7.5.1. *Suppose, in the definition of* naCQP, *that the unitary operators $U_1, \cdots, U_T$ are now allowed to query f. That is, we are given access to the n-qubit gate U_f defined by $U_f|y\rangle = (-1)^{f(y)}|y\rangle$ for all $y \in \{0,1\}^n$, as well as controlled-U_f. Then there is a* naCQP *algorithm to find the value of x that uses $O(N^{1/3})$ queries and $\tilde{O}(N^{1/3})$ time.*

Proof. Prepare the uniform superposition of all basis states, apply $i = N^{1/3}$ Grover iterations [140], then query the oracle to record whether or not each basis state is marked in an

[3]Note that when we say $\mathsf{naCQP}^{\mathcal{O}}$, we mean that circuits given in the input to $\mathcal{Q}$ in the definition of naCQP can contain quantum calls to the oracle.

ancilla. We obtain the state

$$\sin((2i+1)\theta)|x\rangle|1\rangle + \cos((2i+1)\theta) \sum_{y\in\{0,1\}^n, y\neq x} 2^{-\frac{N-1}{2}}|y\rangle|0\rangle$$

where $\sin(\theta) = 2^{-n/2}$ and $i = 2^{n/3}$. For small x we have $\sin(x) \approx x$, so for large n we have $\theta = \Theta(2^{-n/2})$, so $\sin((2i+1)\theta) = \Theta(2^{-n/6})$.

Now make $O(N^{1/3}\log N)$ non-collapsing measurements. We claim that with high probability, the marked item x will appear at least once. Indeed, the marked item x appears with probability at least $\Omega\left(N^{-1/3}\right)$ in each non-collapsing measurement outcome, so it occurs at least once with probability more than $1 - (\log N + 1)e^{-\log N} = 1 - o(1)$. □

Note that if we are willing to use an enormous amount of *time*, we can search in the naCQP model using only one *query*: just query the oracle in superposition and then perform $O(N)$ non-collapsing measurements. Indeed as we note in the introduction, any function f has query complexity 1 in this model, although this approach requires exponentially many non-collapsing measurements. Therefore in this model of computation, the relevant measure of complexity of an algorithm is the number of queries Q plus the number of non-collapsing measurements T used by the algorithm. Our above algorithm uses $Q + T = \tilde{O}(N^{1/3})$ of each, with $O(N^{1/3})$ post-processing time, so we say it "runs in time $\tilde{O}(N^{1/3})$".

7.6 Lower bounds for search

We now show that our search algorithm in section 7.5 cannot be improved by much; in particular there is no way to solve search in faster than $N^{1/4}$ time, even with non-adaptive non-collapsing measurements. Proving the analogous lower bound for adaptive non-collapsing measurements (i.e. for the class CQP) remains open.

Theorem 7.6.1. *Suppose, in the definition of* naCQP*, that the unitary operators $U_1, \cdots, U_T$ are now allowed to query f. Let Q be the number of queries to f made by a* naCQP *algorithm, and T be the number of non-collapsing measurements. Then any* naCQP *algorithm to find the value of x obeys $Q + T = \Omega(N^{1/4})$, and hence search requires $\Omega(N^{1/4})$ time.*

In other words, there is no "black box" polynomial-time algorithm for NP-hard problems, even when given access to non-collapsing measurements. This is evidence that the class naCQP does not contain NP. The following corollary follows immediately from the well-known "diagonalization method" of Baker, Gill, and Solovay [45]:

Corollary 7.6.2. *There exists an oracle $\mathcal{O}$ such that* $\mathsf{NP}^{\mathcal{O}} \not\subset \mathsf{naCQP}^{\mathcal{O}}$.

We now outline the proof of Theorem 7.6.1. The following lemma is essential: it bounds the total variation distance between two Markov distributions.

Lemma 7.6.1. *Suppose that $T \geq 1$, and that $v = (v_0, \cdots, v_T)$ is a random variable governed by a Markov distribution. That is, for all $1 \leq i \leq T$, v_i is independent of $v_0, \cdots, v_{i-2}$ conditioned on a particular value of v_{i-1}. Let $w = (w_0, \cdots, w_T)$ be another random variable governed by a Markov distribution. If $d_{TV}(\cdot,\cdot)$ denotes the total variation distance between random variables, then*

$$d_{TV}(v,w) \leq 2\sum_{i=1}^{T} d_{TV}((v_{i-1}, v_i), (w_{i-1}, w_i)).$$

Proof. We proceed by induction on T. The base case $T = 1$ is trivial. For $T > 1$, since w_T depends only on w_{T-1} (by the Markov property), it is equal to $A(w_{T-1})$ for some randomized process A; let $w'_T := A(v_{T-1})$ be a variable that depends on v_{T-1} in exactly the same way that w_T depends on w_{T-1}. Then, define the random variable $v' = (v_0, \cdots, v_{T-1}, w'_T)$. By the triangle inequality,

$$d_{TV}(v, w) \leq d_{TV}(v, v') + d_{TV}(v', w). \tag{7.1}$$

Applying the same randomized process to two random variables cannot increase their total variation distance [211]. We can generate random variables identically distributed to v and v' by applying a suitable randomized process to (v_{T-1}, v_T) and (v_{T-1}, w'_T). We can also generate random variables identically distributed to v' and w by applying a suitable randomized process to $(v_0, \cdots, v_{T-1})$ and $(w_0, \cdots, w_{T-1})$. Therefore, the right hand side of (7.1) is bounded above by

$$d_{TV}((v_{T-1}, v_T), (v_{T-1}, w'_T)) + d_{TV}((v_0, \cdots, v_{T-1}), (w_0, \cdots, w_{T-1})).$$

By the triangle inequality,

$$\begin{aligned} d_{TV}((v_{T-1}, v_T), (v_{T-1}, w'_T)) &\leq d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)) + d_{TV}((w_{T-1}, w_T), (v_{T-1}, w'_T)) \\ &= d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)) + d_{TV}(v_{T-1}, w_{T-1}) \\ &\leq 2d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)). \end{aligned}$$

Putting all of this together, we get that $d_{TV}(v, w)$ is upper bounded by

$$\begin{aligned} &2d_{TV}((v_{T-1}, v_T), (w_{T-1}, w_T)) \\ &+ d_{TV}((v_0, \cdots, v_{T-1}), (w_0, \cdots, w_{T-1})). \end{aligned}$$

The result follows from induction. □

Lemma 7.6.2. *The trace distance between two pure states $|\psi\rangle$ and $|\phi\rangle$ is less than or equal to the 2-norm $\||\psi\rangle - |\phi\rangle\|_2$.*

Proof. The trace distance between $|\psi\rangle$ and $|\phi\rangle$ is equal to $\sqrt{1 - |\langle\psi|\phi\rangle|^2}$ [199], and the 2-norm $\||\psi\rangle - |\phi\rangle\|_2$ is $\sqrt{2 - 2\mathrm{Re}(\langle\psi|\phi\rangle)}$. The inequality follows from $|\langle\psi|\phi\rangle| \leq 1$. □

From the hybrid argument of [56], we have the following:

Lemma 7.6.3. *For all t, if there are no measurements made before time t, we have*

$$\sum_{x=0}^{N-1} \||\psi_t\rangle - |\psi_t(x)\rangle\|_2^2 \leq 4Q^2.$$

With these facts, we can now prove Theorem 7.6.1. We provide an outline of the proof here, and the full proof can be found in Section 7.10. The basic idea is to realize that the non-collapsing measurement outcomes form a Markov chain, because the distribution of any non-collapsing measurement is independent once the results of the previous intermediate collapsing measurements are fixed. So, letting v and $v(x)$ be the distributions on non-collapsing measurement outcomes when the marked item is absent or present at x, by

applying Lemma 7.6.1, we have that

$$\frac{1}{3} \leq d_{TV}(v, v(x)) \leq 2 \sum_{i=1}^{T} d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x))).$$

Here the lower bound on d_{TV} comes from the fact that our algorithm can distinguish whether or not a marked item is present with probability $2/3$, and hence these distributions must be $1/3$-far apart for all x.

Lemma 7.6.3 tells us that there is some x for which the marginal distributions v_i and $v_i(x)$ are close. However, this isn't sufficient to upper bound the quantity on the right of this inequality, because the correlations between the distributions at steps $i-1$ and i (which are induced by the intermediate collapsing measurements) might make the distributions easier to distinguish[4]. Hence in order to prove this lower bound, we have to substantially strengthen the hybrid argument [56] to show that the correlations induced by the collapsing measurement outcomes do not allow
$d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x)))$ to be large. By carefully keeping track of these induced correlations, we show in Section 7.10 that there is some x for which

$$d_{TV}(v, v(x)) \leq 2 \sum_{i=1}^{T} d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x))) \leq \frac{20TQ}{\sqrt{N}}.$$

where Q is the number of queries made by the algorithm and T is the number of non-collapsing measurements. Combining this with the fact that $d_{TV}(v, v(x)) \geq \frac{1}{3}$ for all x, this implies

$$\frac{20TQ}{\sqrt{N}} \geq \frac{1}{3},$$

and hence the running time of the algorithm is at least $T + Q = \Omega(N^{1/4})$.

7.7 An upper bound on CQP

We now show that CQP is contained in the class $\mathsf{BPP}^{\mathsf{PP}}$. Since naCQP $\subseteq$ CQP, this places both classes in the second level of the counting hierarchy. By comparison, the best known upper bound for BQP is AWPP which is contained in PP [119][24].

Theorem 7.7.1. $\mathsf{CQP} \subseteq \mathsf{BPP}^{\mathsf{PP}}$.

Proof. First note that $\mathsf{BPP}^{\mathsf{PP}} = \mathsf{BPP}^{\#\mathsf{P}}$, because one can always use a PP oracle to count with only polynomial overhead. Therefore it suffices to show $\mathsf{CQP} \subseteq \mathsf{BPP}^{\#\mathsf{P}}$. We now show how to simulate the sampling oracle $\mathcal{Q}$ in $\mathsf{BPP}^{\#P}$. Our algorithm will work for adaptive queries as well. Since $\mathsf{CQP} = \mathsf{BPP}^{\mathcal{Q},1}$, this implies the claim.

Suppose we wish to simulate a sample from the oracle $\mathcal{Q}$ with input circuit
$C = (U_1, M_1, \ldots U_T, M_T)$ on n qubits. Since the choice of gate set does not matter (see Section 7.14), without loss of generality we can assume our circuit is composed of only Toffoli and Hadamard gates, which are universal by a result of Shi [220].

[4]To see how this could happen in general, consider the following two Markov distributions on two bits: D_1 outputs 00 or 11 with equal probability, and D_2 outputs 01 or 10 with equal probability. These have identical marginals on each bit, but are perfectly distinguishable due to the correlations between their bits.

We first simulate the result of the measurement M_1. Suppose without loss of generality that M_1 measures the first k qubits and gets outcome $x_1 \dots x_k \in \{0,1\}^k$. Following the techniques of Adleman, DeMarrais, and Huang [24], we can write the probability that x_1 is 0 or 1 as an exponential sum of poly-time-computable terms (since U_1 is specified by a poly-sized circuit). Since we chose Hadamard and Toffoli as our gate set, all terms in the sum are of the form $\frac{\pm 1}{2^k}$, where k is the number of Hadamard gates in U_1. Hence using the $\#P$ oracle, we can compute $\Pr[x_1 = 1]$ exactly in binary, and then flip a coin with bias p using the base BPP machine to obtain outcome $x_1 \in 0, 1$ with this probability.

We've now sampled the value of x_1. To sample the value of x_2, note that we can also express $\Pr[x_2 = 1 | x_1 = 0]$ as a sum of exponentially many terms, each of which is poly-time computable and takes values in $\frac{\pm 1}{2^k}$. Therefore using the #P oracle, we can exactly compute the *conditional* probability that $x_2 = 1$ given our sampled value of x_1; in other words the #P oracle can compute the probabilities of measurement outcomes under post-selection. In this way we can sample x_2, then x_3, etc. obtain a sample $x_1 \dots x_k \in \{0,1\}^k$ as desired.

Now suppose we wish to sample the variable $v_1 \in \{0,1\}^n$ which is the result of a non-collapsing measurement on the state remaining after measurement M_1 yields value $x_1 \dots x_k$. As noted above, using the #P oracle, we can compute the marginal probability that any qubit is 1, postselected on a particular measurement outcome. Hence using the #P oracle, we can draw the sample v_1 using n queries to the oracle. We can continue this process to simulate M_2, then sample v_2, etc. Therefore we can draw a sample from $\mathcal{Q}$ using $O(nT)$ queries to the #P oracle. Note that this simulation works when the U_i and the M_i are chosen adaptively, since for each t the base BPP machine receives the non-collapsing measurement samples $v_0 \dots v_t$ before proposing the next unitary U_{t+1} and measurement M_{t+1}. Hence this shows $\mathsf{CQP} \subseteq \mathsf{BPP}^{\mathsf{PP}}$.

□

An open question is whether or not we can improve this upper bound to show $\mathsf{naCQP} \subseteq \mathsf{PP}$. This seems difficult because $\mathsf{SZK} \subseteq \mathsf{naCQP}$, and it is open whether or not $\mathsf{SZK} \subseteq \mathsf{PP}$. In fact, inspired by this question, we will give an oracle relative to which SZK is not contained in PP in Chapter 8 of this thesis. Therefore any proof of the containment $\mathsf{CQP} \subseteq \mathsf{PP}$ would require non-relativizing (and in fact, even non-algebrizing [64]) techniques. Note that, as far as we are aware, there are no complexity-theoretic obstructions to improving the upper bound to $\mathsf{P}^{\#\mathsf{P}}$ in a relativizing manner. We leave this as an open question.

7.8 Strange properties of noncollapsing measurements

Here we show why allowing non-collapsing measurements in quantum mechanics allows for faster than light communication, renders quantum query complexity and quantum communication complexity meaningless, and allows for quantum cloning. We also discuss the relationship between naCQP and quantum computers which have the ability to clone.

To see that non-collapsing measurements allow for faster-than-light communication: suppose two players Alice and Bob share n EPR pairs. Then Alice can send a bit of information to Bob with probability $1 - 2^{-n}$. To see this, suppose Alice makes collapsing measurements in the 0/1 basis on her share of the EPR pairs to send a 0, and measures in the +/- basis to send a 1. Now Bob makes two non-collapsing measurements in the 0/1 basis on his half of the EPR pairs. If Alice had sent a 0, then Bob will see the same outcome each time. If Alice had sent a 1, then Bob will see a random string each time, so with probability $1 - 2^{-n}$

Bob will see two different outcomes. Thus Bob can tell which basis Alice measured in with high probability, and Alice and Bob can communicate faster than light.

We now explain why with non-collapsing measurements, the quantum query complexity and quantum communication complexity of any function is 1. Suppose one wishes to evaluate $f(x)$ where $x = x_1 \dots x_N$. Then one can prepare the superposition $\sum_i |i\rangle|x_i\rangle$ with one query to the oracle, and make $O(N \log N)$ non-collapsing measurements of this state to observe the value of each x_i and compute the function. Similarly, in the context of communication complexity, one player can simply encode their input $x \in \{0,1\}^n$ into the state $\cos\theta_x|0\rangle + \sin\theta_x|1\rangle$ where $\theta_x = \frac{x}{2^n}\frac{\pi}{2}$. By performing roughly 2^n non-collapsing measurements, the other player can learn θ_x and hence x, with only one quantum bit of communication. Note that although these example algorithms use only one query or one qubit of communication, respectively, they use a large number of non-collapsing measurements. For this reason, when we prove lower bounds for naCQP, we lower bound the number of queries *plus* the number of non-collapsing measurements required, rather than the number of queries alone.

To see that non-collapsing measurements allow for cloning: given a quantum state ψ on n qubits, one could perform $2^{O(n)}$ non-collapsing measurements to characterize the state using tomography, and then (approximately) reproduce the state. This "approximate cloning" operation takes exponential time for generic states, and is non-unitary.

Note, that the class of computations considered in naCQP cannot clone, even for states of $O(\log(n))$ qubits, since the naCQP machine cannot perform further quantum computations after receiving the non-collapsing measurement results (i.e. because of the non-adaptivity restriction). In contrast, if the circuit could depend on the non-collapsing measurement results (as in CQP), then one could clone states of $O(\log(n))$ qubits to polynomial accuracy, which is a non-unitary operation. Hence following the result of Abrams and Lloyd [22], the power of CQP class might include NP or even #P, though we do not know if this is the case. A broader related open problem is: what is the power of quantum computers which are given the ability to clone? Such devices could clearly simulate naCQP and CQP computations - to simulate a non-collapsing measurement, simply clone and measure in the computational basis. However, it's unclear how powerful such quantum devices would be.

7.9 Missing proof from Section 7.4: a detailed proof that SZK $\subseteq$ naCQP

Here we provide a detailed analysis of the probability of error in our naCQP algorithm for solving the Statistical Difference problem, which is SZK-complete.

Let's briefly recap the algorithm. Suppose we're given an instance of Statistical Difference, and apply the Polarization Lemma of Sahai and Vadhan [211] to obtain two circuits P_0 and P_1 which encode probability distributions D_0 and D_1 which satisfy either $d_{TV}(D_0, D_1) \leq \varepsilon$ or $d_{TV}(D_0, D_1) \geq 1 - \varepsilon$, where $\varepsilon = 2^{-O(n^c)}$ for some constant c. We now prepare the state

$$\frac{1}{2^{(n+1)/2}} \sum_{b \in \{0,1\}, x \in \{0,1\}^n} |b\rangle|x\rangle|P_b(x)\rangle$$

Now, measure the third register with a collapsing measurement to obtain some outcome y, and then perform three non-collapsing measurements on the b register to obtain outcomes b_1, b_2, b_3. If $b_1 = b_2 = b_3$ then output the distributions were $(1 - \varepsilon)$-far in total variation

distance, otherwise output they were ε-close in total variation distance.

We will now compute the probability this algorithm makes an error when the input distributions are ε-close in total variation distance.

Let $D_b(x)$ denote the probability that distribution D_b outputs string x. The probability of seeing outcome y in the $P_b(x)$ register under our collapsing measurement is

$$\frac{1}{2}(D_0(y) + D_1(y)).$$

Conditioned on seeing outcome y, one can easily compute that the probability of obtaining outcome $b_1 = b_2 = b_3$ (which causes the algorithm to err) is $\dfrac{D_0(y)^3 + D_1(y)^3}{(D_0(y) + D_1(y))^3}$.

Hence the total probability of error is this case is given by

$$\Pr[\text{error}] = \sum_y \frac{D_0(y) + D_1(y)}{2} \frac{D_0(y)^3 + D_1(y)^3}{(D_0(y) + D_1(y))^3}$$

Let $\delta(y) = D_1(y) - D_0(y)$. So $\sum_y \delta(y) = 0$ and $\sum_y |\delta(y)| \leq 2\varepsilon$ by our promise on the total variation distance between D_0 and D_1. Hence we have by direct calcuation that

$$\begin{aligned}
\Pr[\text{error}] &= \frac{1}{2} \sum_y \frac{D_0(y)^3 + D_1(y)^3}{(D_0(y) + D_1(y))^2} \\
&= \frac{1}{2} \sum_y \frac{D_0(y)^3 + (D_0(y) + \delta(y))^3}{(2D_0(y) + \delta(y))^2} \\
&= \frac{1}{2} \sum_y \frac{2D_0(y)^3 + 3D_0(y)^2\delta(y) + 3D_0(y)\delta(y)^2 + \delta(y)^3}{4D_0(y)^2 + 4D_0(y)\delta(y) + \delta(y)^2} \\
&= \frac{1}{2} \sum_y \frac{D_0(y)}{2} + \frac{D_0(y)^2\delta(y) + \frac{5}{2}D_0(y)\delta(y)^2 + \delta(y)^3}{4D_0(y)^2 + 4D_0(y)\delta(y) + \delta(y)^2} \\
&= \frac{1}{4} + \frac{1}{2} \sum_y \delta(y) \frac{D_0(y)^2 + \frac{5}{2}D_0(y)\delta(y) + \delta(y)^2}{4D_0(y)^2 + 4D_0(y)\delta(y) + \delta(y)^2} \\
&\leq \frac{1}{4} + \frac{1}{2} \sum_y |\delta(y)| \frac{D_0(y)^2 + \frac{5}{2}D_0(y)|\delta(y)| + |\delta(y)^2|}{4D_0(y)^2 + 4D_0(y)\delta(y) + \delta(y)^2} \\
&\leq \frac{1}{4} + \frac{1}{2} \sum_y |\delta(y)| \frac{4D_0(y)^2 + 4D_0(y)|\delta(y)| + |\delta(y)^2|}{4D_0(y)^2 + 4D_0(y)\delta(y) + \delta(y)^2} \\
&= \frac{1}{4} + \frac{1}{2} \sum_y |\delta(y)| \leq \frac{1}{4} + \varepsilon
\end{aligned}$$

Where the last two lines follow from the fact that all terms in the sum are non-negative, and the fact that $\sum_y |\delta(y)| \leq 2\varepsilon$. Hence the probability of error in the case is upper bounded by $\frac{1}{4} + \varepsilon = \frac{1}{4} + O(2^{-n^c})$, so the algorithm has probability of error $< \frac{1}{3}$ for sufficiently large n as desired.

We now bound the probability of error in the case that the distributions are far apart.

In this case, the probability of getting an outcome where b_1, b_2, b_3 are not all at the same, conditioned on measuring y, is given by

$$\frac{3D_0(y)^2D_1(y) + 3D_0(y)D_1(y)^2}{(D_0(y) + D_1(y))^3}.$$

Hence by direct calculation we have that the probability of error is

$$\begin{aligned}\Pr[\text{error}] &= \sum_y \frac{D_0(y) + D_1(y)}{2} \frac{3D_0(y)^2D_1(y) + 3D_0(y)D_1(y)^2}{(D_0(y) + D_1(y))^3} \\ &= \frac{3}{2}\sum_y \frac{D_0(y)^2D_1(y)}{(D_0(y) + D_1(y))^2} + \frac{D_0(y)D_1(y)^2}{(D_0(y) + D_1(y))^2}\end{aligned}$$

Let us upper bound the first of these terms; the upper bound on the second term follows analogously by switching D_0 and D_1.

Since D_0 and D_1 are $1-\varepsilon$-far in total variation distance, there must exist some set S of y's, and its complement $\bar{S}$, such that $\sum_{y\in S} D_0(y) \geq 1-\varepsilon$ and $\sum_{y\in S} D_1(y) \leq \varepsilon$, which implies that $\sum_{y\in\bar{S}} D_0(y) \leq \varepsilon$ and $\sum_{y\in\bar{S}} \geq 1-\varepsilon$. Hence we have that

$$\begin{aligned}\sum_y \frac{D_0(y)^2D_1(y)}{(D_0(y) + D_1(y))^2} &= \sum_{y\in S} \frac{D_0(y)^2D_1(y)}{(D_0(y) + D_1(y))^2} + \sum_{y\in\bar{S}} \frac{D_0(y)^2D_1(y)}{(D_0(y) + D_1(y))^2} \\ &\leq \sum_{y\in S} \frac{D_0(y)^2D_1(y)}{(D_0(y))^2} + \sum_{y\in\bar{S}} \frac{D_0(y)^2D_1(y)}{(D_1(y))^2} \\ &\leq \sum_{y\in S} D_1(y) + \sum_{y\in\bar{S}} D_0(y) \\ &\leq 2\varepsilon\end{aligned}$$

By applying an analogous bound to the second term, we have that $\Pr[\text{error}] = O(\varepsilon) = O(2^{-n^c})$ as desired, so the probability of error in this case is vanishingly small.

Hence the net probability that the algorithm errs is $\frac{1}{4} + \varepsilon$ in the case the distributions are ε-close and $O(\varepsilon)$ in the case the distributions are $(1-\varepsilon)$-far.

7.10 Missing proof from Section 7.6: An $N^{1/4}$ Lower Bound for Search in naCQP

Here we show that any naCQP algorithm for search requires at least $N^{1/4}$ time.

Proof of Theorem 7.6.1. Since it is always possible to copy measured qubits, we can assume that qubits which are measured in an intermediate step of the algorithm are never directly modified again. Now, assume that the algorithm uses ℓ qubits and applies unitary operators $U_1, \cdots, U_T$, each of which is either a (controlled) query to the search function f or a gate from the finite universal gate set $\mathcal{U}$. The measurements $M_1 \ldots M_T$ (which may or may not be empty) are applied between the operators $U_1 \ldots U_T$.

Let $v(x) = (v_0(x), v_1(x), \cdots, v_T(x))$ be the non-collapsing measurement results when the marked item is x, so that $v_i(x)$ is sampled immediately before the application of U_{i+1}. Let $v = (v_0, \cdots, v_T)$ be the non-collapsing measurement results when there is no marked item. In general, both $v(x)$ and v are random variables. Since the postprocessing step can distinguish the distributions of v and $v(x)$ with success probability $2/3$, $d_{TV}(v, v(x)) \geq \frac{1}{3}$ for all x. On the other hand, each v and $v(x)$ is a Markov process. Therefore, by Lemma 7.6.1,

$$d_{TV}(v, v(x)) \leq 2 \sum_{i=1}^{T} d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x))).$$

Now, we bound the term

$$d_{x,i} := d_{TV}((v_{i-1}, v_i), (v_{i-1}(x), v_i(x))).$$

Since it is possible to defer measurements in a quantum circuit to a later stage [199], we can assume that all intermediate measurements that occurred before the application of U_i occurred immediately before the sampling of v_i. Suppose that these measurements were applied to the first k qubits of the state. Let $|\phi\rangle$ and $|\phi(x)\rangle$ be the state vectors immediately before these measurements. Then, we decompose $|\phi\rangle = \sum_{s \in \{0,1\}^k} \alpha_s |s\rangle |\phi_s\rangle$ and $|\phi(x)\rangle = \sum_{s \in \{0,1\}^k} \beta_s |s\rangle |\phi_s(x)\rangle$. Possible values for (v_{i-1}, v_i) and $(v_{i-1}(x), v_i(x))$ can be written in the form (st_1, st_2), where s is a k-bit string and t_1, t_2 are $(\ell - k)$-bit strings.

Assume for now that U_i does not contain a query to f. Then, since it does not affect the first k qubits, it can be decomposed into the sum $\sum_{s \in \{0,1\}^k} |s\rangle V_s \langle s|$ for some unitary operators V_s. The transformation U_i can be thought of as applying the unitary V_s to the last $\ell - k$ qubits if the (measured) first k qubits are equal to s. Then, the probability that $(v_{i-1}, v_i) = (st_1, st_2)$ is equal to $|\alpha_s|^2 |\langle t_1 | \phi_s \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2$, and the probability that $(v_{i-1}(x), v_i(x)) = (st_1, st_2)$ is equal to

$$|\beta_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2.$$

Therefore, the total variation distance $d_{x,i}$ is by the triangle inequality

$$\begin{aligned}
d_{x,i} &= \frac{1}{2} \sum_{s,t_1,t_2} \left| |\alpha_s|^2 |\langle t_1 | \phi_s \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 - |\beta_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right| \\
&\leq \frac{1}{2} \sum_{s,t_1,t_2} \left(\left| |\alpha_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 - |\beta_s|^2 |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right| \right) \\
&\quad + \frac{1}{2} \sum_{s,t_1,t_2} \left(|\alpha_s|^2 \left| |\langle t_1 | \phi_s \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 - |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 \right| \right) \\
&\quad + \frac{1}{2} \sum_{s,t_1,t_2} \left(|\alpha_s|^2 \left| |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s \rangle|^2 - |\langle t_1 | \phi_s(x) \rangle|^2 |\langle t_2 | V_s | \phi_s(x) \rangle|^2 \right| \right) \\
&=: \frac{1}{2}(S_1 + S_2 + S_3)
\end{aligned}$$

where S_1, S_2, S_3 are the three sums written above, which range over $s \in \{0,1\}^k$ and $t_1, t_2 \in \{0,1\}^{\ell-k}$. Now, we have:

$$\begin{aligned}
S_1 &:= \sum_{s,t_1,t_2} \left(||\alpha_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2|V_s|\phi_s(x)\rangle|^2 - |\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2|V_s|\phi_s(x)\rangle|^2 | \right) \\
&= \sum_s \left| |\alpha_s|^2 - |\beta_s|^2 \right| \left(\sum_{t_1,t_2} |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2|V_s|\phi_s(x)\rangle|^2 \right) \\
&= \sum_s \left| |\alpha_s|^2 - |\beta_s|^2 \right| \\
&\leq \| |\phi\rangle\langle\phi| - |\phi(x)\rangle\langle\phi(x)| \|_{tr} \\
&\leq 2 \| |\phi(x)\rangle - |\phi\rangle \|_2 .
\end{aligned}$$

Additionally,

$$\begin{aligned}
S_2 &:= \sum_{s,t_1,t_2} \left(|\alpha_s|^2 | |\langle t_1|\phi_s\rangle|^2 |\langle t_2|V_s|\phi_s\rangle|^2 |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2|V_s|\phi_s\rangle|^2 | \right) \\
&= \sum_{s,t_1} \left(|\alpha_s|^2 \left| |\langle t_1|\phi_s\rangle|^2 - |\langle t_1|\phi_s(x)\rangle|^2 \right| \right) \\
&\leq \sum_{s,t_1} \left(\left| |\alpha_s|^2 |\langle t_1|\phi_s\rangle|^2 - |\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 \right| \right) \\
&\quad + \sum_{s,t_1} \left(\left| |\alpha_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 - |\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 \right| \right) \\
&= \sum_{s,t_1} \left(\left| |\alpha_s|^2 |\langle t_1|\phi_s\rangle|^2 - |\beta_s|^2 |\langle t_1|\phi_s(x)\rangle|^2 \right| \right) + \sum_s \left(\left| |\alpha_s|^2 - |\beta_s|^2 \right| \right) \\
&\leq 2 \| |\phi\rangle\langle\phi| - |\phi(x)\rangle\langle\phi(x)| \|_{tr} \\
&\leq 4 \| |\phi(x)\rangle - |\phi\rangle \|_2 .
\end{aligned}$$

Finally,

$$\begin{aligned}
S_3 &= \sum_{s,t_1,t_2} \left(|\alpha_s|^2 | |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2|V_s|\phi_s\rangle|^2 - |\langle t_1|\phi_s(x)\rangle|^2 |\langle t_2|V_s|\phi_s(x)\rangle|^2 | \right) \\
&= \sum_{s,t_2} \left(|\alpha_s|^2 \left| |\langle t_2|V_s|\phi_s\rangle|^2 - |\langle t_2|V_s|\phi_s(x)\rangle|^2 \right| \right) \\
&\leq \sum_{s,t_2} \left(\left| |\alpha_s|^2 |\langle t_2|V_s|\phi_s\rangle|^2 - |\beta_s|^2 |\langle t_2|V_s|\phi_s(x)\rangle|^2 \right| \right) \\
&+ \sum_{s,t_2} \left(\left| |\alpha_s|^2 |\langle t_2|V_s|\phi_s(x)\rangle|^2 - |\beta_s|^2 |\langle t_2|V_s|\phi_s(x)\rangle|^2 \right| \right) \\
&= \sum_{s,t_2} \left(\left| |\alpha_s|^2 |\langle t_2|V_s|\phi_s\rangle|^2 - |\beta_s|^2 |\langle t_2|V_s|\phi_s(x)\rangle|^2 \right| \right) + \sum_s \left(\left| |\alpha_s|^2 - |\beta_s|^2 \right| \right) \\
&\leq 2 \| |\phi\rangle\langle\phi| - |\phi(x)\rangle\langle\phi(x)| \|_{tr} \\
&= 4 \| |\phi(x)\rangle - |\phi\rangle \|_2
\end{aligned}$$

Therefore,

$$d_{x,i} \leq \frac{1}{2}(S_1 + S_2 + S_3) \leq 5 \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2 .$$

On the other hand, if U_i is a query to f, then it only applies a local phase of -1 to some of the probability amplitudes of $|\phi\rangle$ and $|\phi_x\rangle$. Therefore, the same argument still shows that $d_{x,i} \leq 5 \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2$.

By the Cauchy-Schwarz inequality and Lemma 7.6.3,

$$\begin{aligned} \frac{1}{N}\sum_{x=0}^{N-1} d_{x,i} &\leq 5 \cdot \frac{1}{N}\sum_{x=0}^{N-1} \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2 \\ &\leq 5\sqrt{\frac{1}{N}\sum_{x=0}^{N-1} \left\| |\phi(x)\rangle - |\phi\rangle \right\|_2^2} \\ &\leq \frac{10Q}{\sqrt{N}} \end{aligned}$$

for all i. Therefore, there is some x for which

$$d_{TV}(v, v(x)) \leq 2\sum_{i=1}^{T} d_{x,i} \leq \frac{20TQ}{\sqrt{N}}.$$

On the other hand, $d_{TV}(v, v(x)) \geq \frac{1}{3}$ for all x, so

$$\frac{20TQ}{\sqrt{N}} \geq \frac{1}{3},$$

and the running time of the algorithm is at least $T + Q = \Omega(N^{1/4})$. □

7.11 An $N^{1/3}$ lower bound for search in naCQP if there are no collapsing measurements

Assume that intermediate measurements are not allowed in our search algorithm. As we discuss in Section 7.3, this gives a model with only the power of BQP, because then the states $|\psi_t\rangle = U_t U_{t-1} \ldots U_1 |0\rangle^{\otimes n}$ can be generated with poly-sized circuits, and hence a BQP machine could prepare and and measure them to sample from $\mathcal{Q}$. Trivially one can prove a lower bound of $N^{1/4}$ for search in this model, either by noting that this class can achieve at most quadratic speedups over BQP by the previous comment, or by using the argument put forth in Theorem 7.6.1. Here we tighten this result to give an $N^{1/3}$ lower bound for search in this class.

Suppose that an algorithm A searches with Q queries and T timesteps, where $Q + T = o(N^{1/3})$. Let ψ_t be the quantum state after t steps with no marked item, and let ψ_t^x be defined likewise when the marked item is at location x. By the hybrid argument we have that $\forall t$

$$\sum_x ||\psi_t - \psi_t^x||_2^2 \leq 4Q^2$$

where $||a||_2^2$ is the 2-norm squared of a. This implies

$$\sum_t \sum_x ||\psi_t - \psi_t^x||_2^2 \leq 4TQ^2$$

Hence there must exist x such that

$$\sum_t ||\psi_t - \psi_t^x||_2^2 \leq \frac{4TQ^2}{N} \tag{7.2}$$

Since we assumed $Q + T = o(N^{1/3})$, we have that $\frac{4TQ^2}{N} = o(1)$. Therefore for sufficiently large N and for all t we have

$$||\psi_t - \psi_t^x||_2^2 \leq 0.01$$

(The choice of constant here is arbitrary, we simply need it to be less than around 0.5.) Now consider the states $\Psi := \bigotimes_t |\psi_t\rangle$ and $\Psi^x := \bigotimes_t |\psi_t^x\rangle$. Let V the distribution on samples with no marked item, and let V^x be defined likewise. Then clearly we have that

$$|V - V^x|_1 \leq ||\Psi - \Psi^x||$$

where $||a||$ denotes the trace norm of a. This is because the output distributions of V and V^x can be obtained by (independent) measurements on the states Ψ and Ψ_x in the computational basis. Note that $|V - V^x|_1$ must be $\Omega(1)$ in order to distinguish the presence of a marked item at x in postprocessing. Therefore we have

$$\begin{aligned}
\Omega(1) \leq |V - V_x|_1 &\leq ||\Psi - \Psi^x|| && (7.3)\\
&= \sqrt{1 - |\langle\Psi|\Psi^x\rangle|^2} && (7.4)\\
&= \sqrt{1 - |\Pi_t \langle\psi_t|\psi_t^x\rangle|^2} && (7.5)\\
&\leq \sqrt{1 - \Pi_t e^{-||\psi_t - \psi_t^x||_2^2}} && (7.6)\\
&= \sqrt{1 - e^{-\Sigma_t ||\psi_t - \psi_t^x||_2^2}} && (7.7)\\
&\leq \sqrt{1 - e^{-\frac{4TQ^2}{N}}} && (7.8)\\
&= o(1) && (7.9)
\end{aligned}$$

Where in line 7.4 we use the formula for trace distance of pure states, in line 7.8 we used equation 7.2, in line 7.9 we used the fact that $T + Q = o(N^{1/3})$, and in line 7.6 we use the inequality

$$\begin{aligned}
|\langle\psi_t|\psi_t^x\rangle| &\geq \mathrm{Re}\left(\langle\psi_t|\psi_t^x\rangle\right) && (7.10)\\
&= 1 - \frac{||\psi_t - \psi_t^x||_2^2}{2} && (7.11)\\
&\geq e^{-||\psi_t - \psi_t^x||_2^2} && (7.12)
\end{aligned}$$

where we have use the fact that $1 - x \geq e^{-2x}$ for $0 \leq x \leq 0.01$. Therefore we have shown $\Omega(1) = o(1)$, a contradiction. Hence such an algorithm A cannot exist, so searching takes

$Q + T = \Omega(N^{1/3})$ time when there are non-collapsing measurements, but no collapsing measurements, in the model.

7.12 The error in the DQP search time lower bound

We now describe the error in Aaronson's original proof of an $\Omega(N^{1/3})$ lower bound for search in the DQP model, which is related to the fact that hidden variable theories can have strong correlations between their values at different times.

7.12.1 The class DQP

We first describe the formal definition of the complexity class DQP, which is based on the notion of a hidden-variable theory. A hidden-variable theory is an interpretation of quantum mechanics in which a quantum system is described by both a state vector and a definite state (called the "hidden variable"), which determines the result of measurements on the system. When a transformation is applied to the system, the state vector evolves by a unitary linear transformation, like in ordinary quantum mechanics, and the hidden variable evolves stochastically according to the state vector and the unitary linear transformation. According to the Kochen-Specker theorem [205], it is impossible for the hidden variable to determine a result for all possible measurements on the system. Therefore, in what follows, we will only ever measure the quantum system in some fixed basis.

Suppose that our quantum system is described by a Hilbert space with N basis states $|1\rangle, |2\rangle, \ldots, |N\rangle$. Then, the hidden variable has one of the values $1, \ldots, N$. The hidden-variable theory specifies the probabilities that the hidden variable changes from i to j given that the state was $|\psi\rangle$ and was transformed by the unitary U. More precisely, a hidden variable theory $\mathcal{T}$ is specified by a stochastic matrix $S_{\mathcal{T}}(|\psi\rangle, U)$ for every state $|\psi\rangle$ and unitary transformation U of dimension N, which indicates how the hidden variable evolves when the state transforms from $|\psi\rangle$ to $U|\psi\rangle$. If $\mathcal{T}$ is understood from context, then we simply write $S(|\psi\rangle, U)$. Suppose $|\psi\rangle = \Sigma_i \alpha_i |i\rangle$ and $U|\psi\rangle = \Sigma_j \beta_j |j\rangle$. The hidden-variable theory must be consistent with the predictions of quantum mechanics, which is to say that the probability that the hidden variable is equal to i is equal to $|\alpha_i|^2$. This means that the stochastic matrix $S = S(|\psi\rangle, U)$ must satisfy

$$|\beta_j|^2 = \sum_{i=1}^{n} |\alpha_i|^2 (S)_{ij}.$$

Other "reasonable" properties that we might expect a hidden-variable theory to have, for example that $S(|\psi\rangle, WV) = S(|\psi\rangle, V) S(V|\psi\rangle, W)$, need not be satisfied.

Sometimes, the hidden-variable theory is described instead by the matrix $P = P(|\psi\rangle, U)$ of joint probabilities, defined by $(P)_{ij} = |\alpha_i|^2 (S)_{ij}$. The matrix S is then recovered by

$$S(|\psi\rangle, U) = \lim_{\varepsilon \to 0^+} \frac{(P(|\psi_\varepsilon\rangle, U))_{ij}}{|(|\psi_\varepsilon\rangle)_i|^2}$$

where $|\psi_\varepsilon\rangle = \sqrt{1-\varepsilon}|\psi\rangle + \sqrt{\varepsilon}\frac{1}{2^{N/2}}\Sigma_i |i\rangle$. The function $P(|\psi\rangle, U)$ only defines a hidden-variable theory if this limit actually exists.

The hidden-variable theory is called *local* if unitary transformations on some subsystem A of the system do not affect the value of the hidden variable on a separate subsystem B.

A stronger property is *indifference*, which is the property that if U is block-diagonal, then $S(|\psi\rangle, U)$ is block-diagonal with the same block structure or some refinement thereof. It is called *commutative* if the order of unitaries applied to separate subsystems is irrelevant. A theorem of Bell states that no hidden-variable theory satisfies both locality and commutativity. The theory is called *robust* if for every polynomial $q(N)$, there is a polynomial $p(N)$ such that perturbing the unitary U and density matrix $|\psi\rangle$ by at most $\frac{1}{p(N)}$ in the infinity norm changes the matrix $P(|\psi\rangle, U)$ by at most $\frac{1}{q(N)}$ in the infinity norm. An example of a robust indifferent hidden variable theory is the flow theory $\mathcal{FT}$ defined in [5], which is based on network flows. For a more detailed treatment of hidden variable theories, see [5].

The complexity class DQP (Dynamical Quantum Polynomial Time) is the class of all problems solvable efficiently in the dynamic quantum model of computation. The basic idea is that a dynamic quantum algorithm is allowed to see the whole history of a hidden variable through some quantum computation (and postprocess it classically), as opposed to a quantum algorithm which can only see the final value of the hidden variable.

More formally, suppose that $U_1, \cdots, U_T$ are unitary transformations on ℓ qubits, each specified by a sequence of gates from some finite universal gate set $\mathcal{U}$. Then, a *history* of the hidden variable is a sequence $(v_0, \cdots, v_T)$ of computational basis states, with $v_0 = |0\rangle^{\otimes \ell}$. For any hidden-variable theory $\mathcal{T}$, the rule

$$\Pr[v = (v_0, \cdots, v_T)] = \prod_{k=0}^{T-1} (S_{\mathcal{T}}(U_k \cdots U_1 |0\rangle^{\otimes \ell}, U_{k+1}))_{v_k v_{k+1}}$$

defines a Markov distribution on histories. The oracle $\mathcal{O}(\mathcal{T})$ takes as input the unitaries $(U_1, U_2, \cdots, U_T)$, specified by sequences of gates from $\mathcal{U}$, and outputs a sample from this distribution.

Now, we are ready to define the complexity class DQP. The computational model is a deterministic classical polynomial-time Turing machine A that is allowed one oracle query to $\mathcal{O}(\mathcal{T})$. A language L is in DQP if there is such a Turing machine A, such that for *any* robust indifferent hidden-variable theory $\mathcal{T}$, the machine A correctly decides, with probability at least $2/3$, whether a string of length n is in L, for all sufficiently large n. It follows from the principle of deferred measurement that DQP $\supset$ BQP, because viewing the entire history of a quantum system is at least as powerful as observing it only at the end of a computation [5]. It is important that there is one machine A that works for all robust indifferent hidden-variable theories $\mathcal{T}$.

7.12.2 The error

We now describe the error in Aaronson's proof that any algorithm for the search problem in DQP takes at least $\Omega(N^{1/3})$ time. His proof is based on the hybrid argument: it shows that changing the marked item from x to x^* does not affect the distribution of any particular entry v_i of the hidden-variable history by very much (in the total variation distance). This part of the proof is correct. However, from there he claims that this implies the total variation distance between the entire hidden variable histories v, w is small, using the following inequality

$$d_{TV}(v,w) \leq \sum_{i=0}^{T} d_{TV}(v_i, w_i).$$

While this inequality looks quite similar to Lemma 7.6.1, it is false. The reason is that correlations between the v_i's in a Markov chain can cause the total variation distance between the Markov chains to be high, while the total variation distance between the marginals is small. A specific counterexample is $T = 1$, where v is $(0,0)$ with probability $\frac{1}{2}$ and $(1,1)$ with probabilty $\frac{1}{2}$, and w is $(0,1)$ with probability $\frac{1}{2}$ and $(1,0)$ with probabilty $\frac{1}{2}$. These distributions are perfectly distinguishable, but they have the property that their marginals on any entry are identical (a 50-50 coin flip). Hence

$$d_{TV}(v,w) = 1$$

for this distribution whereas

$$\sum_{i=0}^{T} d_{TV}(v_i, w_i) = 0$$

Although $d_{TV}(v,w)$ cannot be upper bounded in this way, this sort of argument does show that for some item location, the probability of seeing the marked item in the hidden variable history is upper bounded by $O\left(\frac{Q^2T}{N}\right)$ (this follows from the hybrid argument and the union bound). So any search algorithm in DQP which is required to see the marked item takes at least $\Omega(N^{1/3})$ time. However, it is possible that a DQP algorithm could infer the marked item's presence by observing correlations in the hidden variable history, without ever seeing the marked item itself. This possibility is what breaks the proof.

In order to fix this step in Aaronson's proof, one would have to show that the quantity $d_{TV}((v_{i-1}, v_i), (w_{i-1}, w_i))$ is small for each i, and then apply Lemma 7.6.1 to bound the total variation distance between v and w. Furthermore, since a DQP algorithm is required to work for all indifferent or robust hidden variable theories, one would only need to exhibit a single hidden variable theory in which is this is small. However, we only know of one indifferent and robust hidden variable theory ("flow theory"), and it remains open whether or not it satisfies this property.

7.12.3 A proposed roadmap for fixing the error

One way to fix this lower bound would be to find a hidden variable theory which is extremely robust to small perturbations. By the hybrid argument, we know that for any search algorithm making few queries, there will exist a marked item x for which the state of the system $|\psi^x\rangle$ with the item x present is ε-close (where $\varepsilon \approxeq \frac{Q}{\sqrt{N}}$) to the state $|\psi\rangle$ without the marked item.

Call a hidden variable theory *strongly robust* if, for all states ψ, ϕ that are ε-close, and all U, U' that are ε-close,

$$|P(\psi, U) - P(\phi, U')|_1 \leq \text{poly}(\varepsilon)\,\text{polylog}(N)$$

In other words, perturbing the states only perturbs the joint probability matrices by a small

amount, which increases only polynomially in the number of qubits. In contrast, a robust theory is only required to obey $|P(\psi,U) - P(\phi,U')|_1 \leq \text{poly}(\varepsilon)\,\text{poly}(N)$, i.e. the joint probability matrices can be perturbed by an amount which increases polynomially in the dimension of the Hilbert space.

If a strongly robust a theory exists, it would immediately imply a lower bound for search in DQP which is polynomial in N - the reason is that for this marked item x, we would have

$$|P(\psi,U) - P(\psi^x,U^x)|_1 \leq \text{poly}(\varepsilon)\,\text{polylog}(N) = \text{poly}\left(\frac{Q}{\sqrt{N}}\right)\text{polylog}(N)$$

at all stages of the algorithm, and hence by Lemma 7.6.1,

$$\begin{aligned} d_{TV}(v,v^x) &\leq \sum_i d_{TV}((v_{i-1},v_i),(v^x_{i-1},v^x_i)) \\ &= \sum_t |P(\psi^x_t,U^x_t) - P(\psi_t,U_t)|_1 \\ &\leq T\,\text{poly}\left(\frac{Q}{\sqrt{N}}\right)\text{polylog}(N) \end{aligned}$$

Since the DQP search algorithm must work for this strongly robust theory, we must have

$$\frac{TQ^c\,\text{polylog}(N)}{N^{c/2}} \geq d_{TV}(v,v^x) \geq \Omega(1)$$

for some constant c which is the exponent of the polynomial in ε. This implies $T + Q = \tilde{\Omega}(N^{c/(2+2c)})$. Note that perturbing a state by ε has to perturb the resulting P matrices by at least ε (since it must alter their row sums by ε), and hence we must have $0 < c \leq 1$. Therefore even if a strongly robust theory exists, the best possible lower bound one could prove using this technique is $N^{1/4}$.

Unfortunately we do not know of any theories which are strongly robust. The only provably robust theory we know of is flow theory, which in [5] is shown to obey

$$|P(\psi,U) - P(\psi^x,U^x)|_1 \leq 4\varepsilon N^2$$

which does not meet the criteria for strong robustness. An interesting open problem is to determine if flow theory, Schrodinger theory (described in [5]), or any hidden variable theory is strongly robust.

7.13 An $N^{1/4}$ lower bound for search in a modified version of DQP

Although we do not know how to prove a polynomial lower bound for search in DQP, we can show an $N^{1/4}$ lower bound for search in a modified version of DQP, which we describe below:

We first modify the definition of a hidden variable theory. A hidden variable theory is a function $P(\psi,C)$ which depends on

1. A quantum state $\psi = \Sigma_i \alpha_i |i\rangle$
2. A quantum circuit C which specifies product of unitary gate elements g^k, where $k =$

$1 \ldots poly(n)$, from some universal gate set $\mathcal{U}$. Note $U = \Pi_k g^k$.

Unlike before, we now allow $P(\psi, C)$ to depend on the circuit generating the unitary U, rather than only the unitary itself. The output of $P(\psi, C)$ is a joint probability matrix P_{ij}, $i, j = 1 \ldots N$ which satisfies

1. $\Sigma_j P_{ij} = |\alpha_i|^2$ where $\psi = \Sigma_i \alpha_i |i\rangle$
2. $\Sigma_i P_{ij} = |\beta_j|^2$ where we have $U\psi = \Sigma_j \beta_j |j\rangle$

as before.

We call $B \subseteq [N]$ a *circuit block* for circuit $C = \Pi_k g^k$, where each g^k is a gate from a universal gate set $\mathcal{U}$, if for all k, $g^k_{ij} = 0$ for all $i \in B, j \notin B$ and $g^k_{ij} = 0$ for all $i \notin B$, $j \in B$. In other words, a circuit block B is valid if for all circuit elements g_k, indices i, j are in the same block in the unitary g_k. The *circuit block structure* of C is a minimal collection of circuit blocks which partition $[N]$.

In contrast, the block structure of C is the block structure of the resulting unitary. Note that block structure of C is always a refinement of its circuit block structure; if all gates in C have B as a valid block, then the final unitary will have B as a valid block, but the converse is not true. For example, suppose that $C = HH$ on a single qubit. Since $U = HH = I$ the block structure of C is $\{1\}, \{2\}$. However the circuit block structure of C is $\{1, 2\}$, i.e. the trivial circuit block structure, because the individual circuit elements do not have any block structure.

We call a hidden variable theory *circuit-indifferent* if $P(\psi, C)$'s block structure respects the circuit block structure of C. Since the block structure of a unitary U is always a refinement of the circuit block structure of the circuit C producing U, an indifferent theory is always circuit-indifferent. Hence the set of circuit-indifferent theories is larger than the set of indifferent theories.

We define a new version of DQP, which we call CDQP (for "circuit-indifferent DQP"), as before, except

1. We require the algorithms to work for all circuit-indifferent hidden variable theories
2. We no longer require the hidden variable theories to be robust. As a result the definition of our class is gate set dependent. Assume we have all 1 and 2-qubit gates at our disposal.
3. When given access to a search oracle $f : \{0,1\}^n \to \{0,1\}$, we assume it is a phase oracle, i.e. $\mathcal{O}_f |x\rangle = (-1)^{f(x)} |i\rangle$. This distinction did not matter in the definition of DQP or naCQP, but it does matter here, because our hidden variable theories depend on the block structure of individual circuit elements, including the oracle.

We can now prove a lower bound for search in this version of CDQP.

Theorem 7.13.1. *Any algorithm correctly deciding search in* CDQP *using Q queries and T time satisfies $Q + T = \Omega(N^{1/4})$.*

Proof. We will describe a circuit-indifferent hidden variable theory, which we call Dieks theory for circuit block structure, which foils any search algorithm A which uses $Q + T = o(N^{1/4})$ time. This contradicts the requirement that A work for all circuit-indifferent hidden variable theories.

Suppose that A generates quantum circuits $C_1 \ldots C_T$ when there is no marked item, and quantum circuits $C_1^x \ldots C_T^x$ when there is a marked item at location x. Clearly the circuits C_t and C_t^x differ only in their search oracles. The search oracles are diagonal, hence C_t and C_t^x have the same circuit block structure I. This will be crucial in proving our result.

Let ψ_t be the quantum state after t steps of the algorithm when there is no marked item, and let ψ_t^x be the quantum state after t steps when there is a marked item at location x. By the hybrid argument, there exists an item x such that

$$||\psi_t - \psi_t^x|| \leq \frac{4Q}{\sqrt{N}} \tag{7.13}$$

for all $t = 1 \ldots T$, where $||\psi_t - \psi_t^x||$ indicates the trace norm.

We will show that if $P(\psi_t, C_{t+1}, t)$ and $P(\psi_t^x, C_{t+1}^x)$ are given by Dieks theory for circuit block structure, then

$$|P(\psi_t, C_{t+1}) - P(\psi_t^x, C_{t+1}^x)|_1 \leq \frac{12Q}{\sqrt{N}} \tag{7.14}$$

From this the lower bound will follow, because the trace distance between the hidden variable histories with and without a marked item is upper bounded by

$$\sum_t |P(\psi_t, C_{t+1}) - P(\psi_t^x, C_{t+1}^x)|_1 \leq O\left(\frac{TQ}{\sqrt{N}}\right)$$

by Lemma 7.6.1. The quantity must be $\Omega(1)$ because A distinguishes the presense of a marked item with $\Omega(1)$ probability. Hence we have $TQ = \Omega(N^{1/2})$ so $T + Q = \Omega(N^{1/4})$ as desired.

We now define Dieks theory for circuit block structure. Let I be the circuit block structure of C. Let $P := P(\psi_t, C_{t+1})$ be the joint probability matrix of Dieks theory with block structure I. That is,

$$P_{ij} = |\alpha_i|^2 \frac{|\beta_j|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2}$$

if i, j are in the same block $B \in I$ and 0 otherwise. Note P is a valid, circuit indifferent matrix. Indeed the column and row sums are

$$\sum_j P_{ij} = |\alpha_i|^2 \sum_{j \in B} \frac{|\beta_j|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2} = |\alpha_i|^2 \tag{7.15}$$

$$\sum_i P_{ij} = \sum_{i \in B} |\alpha_i|^2 \frac{|\beta_j|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2} \tag{7.16}$$

$$= |\beta_j|^2 \frac{\Sigma_{i \in B} |\alpha_i|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2} = |\beta_j^2| \tag{7.17}$$

where in line 7.17 we used the fact that the actual block structure of U is a refinement of the circuit block structure of C, hence U restricted to any block B of I is also unitary, and so $\Sigma_{i \in B} |\alpha_i|^2 = \Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2$. Hence $P(\psi, C)$ is a valid circuit-indifferent hidden variable theory.

The following Lemma, combined with the equation 7.13 and the fact that C and C^x have the same circuit block structure, implies equation 7.14.

Lemma 7.13.1. *Suppose that* $||\psi - \psi_x|| \leq ||U\psi - U^x\psi^x|| \leq \varepsilon$ *where* U (U^x) *is the unitary produced by circuit* C (C^x). *Furthermore suppose* C *and* C^x *have the same circuit block structure. Then if* P *is given by Dieks theory for circuit block structure, then* $|P(\psi, C) - P(\psi^x, C^x)| \leq 3\varepsilon$.

Proof. Let $\alpha_i, \alpha_i^x, \beta_i, \beta_i^x$ be defined by $\psi = \Sigma_i \alpha_i |i\rangle$, $\psi^x = \Sigma_i \alpha_i^x |i\rangle$, $U\psi = \Sigma_i \beta_i |i\rangle$, and $U\psi^x = \Sigma_i \beta_i^x |i\rangle$ as usual.

Let I be the circuit block structure of C and C^x. By the definition of Dieks theory for circuit indifference, we have that $P := P(\psi, C)$ and $\hat{P} := P(\psi^x, C^x)$ are given by

$$P_{ij} = \begin{cases} |\alpha_i|^2 \frac{|\beta_j|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2} & i, j \in B \in I \\ 0 & \text{o.w.} \end{cases} \qquad \hat{P}_{ij} = \begin{cases} |\alpha_i^x|^2 \frac{|\beta_j^x|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}^x|^2} & i, j \in B \in I \\ 0 & \text{o.w.} \end{cases}$$

We can now show $\hat{P}$ is close to P in trace distance. Note that

$$|P - \hat{P}|_1 = \sum_{i,j} \left| |\alpha_i|^2 \frac{|\beta_j|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2} - |\alpha_i^x|^2 \frac{|\beta_j^x|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}^x|^2} \right| \tag{7.18}$$

$$\leq \sum_B \sum_{i,j \in B} \left| |\alpha_i|^2 \frac{|\beta_j|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2} - |\alpha_i|^2 \frac{|\beta_j^x|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}^x|^2} \right| + \left| |\alpha_i|^2 \frac{|\beta_j^x|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}^x|^2} - |\alpha_i^x|^2 \frac{|\beta_j^x|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}^x|^2} \right| \tag{7.19}$$

$$= \sum_B \sum_{i,j \in B} |\alpha_i|^2 \left| \frac{|\beta_j|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2} - \frac{|\beta_j^x|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}^x|^2} \right| + \sum_B \sum_{i,j \in B} \frac{|\beta_j^x|^2}{\Sigma_{\hat{j} \in B} \Sigma_{i \in B} |\beta_{\hat{j}}^x|^2} \left| |\alpha_i|^2 - |\alpha_i^x|^2 \right| \tag{7.20}$$

$$= \sum_B \sum_{j \in B} \sum_{\hat{j} \in B} |\beta_{\hat{j}}|^2 \left| \frac{|\beta_j|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2} - \frac{|\beta_j^x|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}^x|^2} \right| + \sum_i \left| |\alpha_i|^2 - |\alpha_i^x|^2 \right| \tag{7.21}$$

$$\leq \sum_B \sum_{j \in B} \left| |\beta_j|^2 - |\beta_j^x|^2 \frac{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}^x|^2} \right| + \varepsilon \tag{7.22}$$

$$\leq \sum_B \sum_{j \in B} \left| |\beta_j|^2 - |\beta_j^x|^2 \right| + \left| |\beta_j^x|^2 - |\beta_j^x|^2 \frac{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}^x|^2} \right| + \varepsilon \tag{7.23}$$

$$\leq \varepsilon + \sum_B \sum_{j \in B} |\beta_j^x|^2 \left| 1 - \frac{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}|^2}{\Sigma_{\hat{j} \in B} |\beta_{\hat{j}}^x|^2} \right| + \varepsilon \tag{7.24}$$

$$= \varepsilon + \sum_B \left| \sum_{j \in B} |\beta_j^x|^2 - \sum_{\hat{j} \in B} |\beta_{\hat{j}}|^2 \right| + \varepsilon \tag{7.25}$$

$$\leq 3\varepsilon \tag{7.26}$$

where line (7.19) follows from the triangle inequality, line (7.21) from the fact that U has

block structure I so $\Sigma_{i\in B}|\alpha_i|^2 = \Sigma_{j\in B}|\beta_j|^2$ as well as an evaluation of the second sum, line (7.22) from our upper bound on the trace distance of ψ and ψ_x, line (7.23) by the triangle inequality, and lines (7.24) and (7.26) by our upper bound on the trace distance of $U\psi$ and $U\psi^x$. This completes the proof.

□

Hence Dieks theory for circuit block structure foils any CDQP algorithm taking less than $N^{1/4}$ time, which completes the proof.

□

7.14 Universal gate set does not matter

We prove that the universal gate set $\mathcal{U}$ used in the definition of naCQP does not matter (a similar proof holds for CQP). Our proof relies on Lemma 7.6.1 and the Solovay-Kitaev theorem [99] to show that any computation using a particular universal gate set $\mathcal{U}$ can be done using a different gate set $\mathcal{U}'$ in such a way that the distributions of the histories does not change significantly in total variation distance.

To do so, we will first give an alternative definition of naCQP which will make the proof easier. Our alternative definition is framed in the notation of DQP; for an introduction to this notation please see Section 7.12.

7.14.1 An alternative definition of naCQP

If B is a partition of $\{0,1\}^\ell$ and U is a unitary operator on $(\mathbb{C}^2)^{\otimes\ell}$, then we say that U respects the block structure B if $U_{ij} = 0$ whenever i and j are in different parts of B. If $|\psi\rangle$ is a pure state and U is a unitary that respects the block structure B, then the stochastic matrix $S_{\mathcal{PT}_B}(|\psi\rangle, U)$ is formed by applying the "product theory" $\mathcal{PT}$ separately on each block of B. More precisely, let $\sim$ be the equivalence relation on $\{1,\cdots,n\}$ defined by $i \sim j$ if and only if i and j are in the same block of B. Let $|\psi\rangle = \Sigma_i \alpha_i |i\rangle$ and $U|\psi\rangle = \Sigma_j \beta_j |j\rangle$. Then,

$$(S_{\mathcal{PT}_B}(|\psi\rangle, U))_{ij} = \begin{cases} \dfrac{|\beta_j|^2}{\sum_{k\sim j}|\beta_k|^2} & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

where the sum over k ranges over all k with $k \sim j$.

Suppose that $\mathcal{V} = (U_1, \cdots, U_T)$ are unitary operators on ℓ qubits, and $\mathcal{B} = (B_1, \cdots, B_T)$ are partitions of $\{0,\cdots,1\}^n$ such that for every i, B_{i+1} is a refinement of B_i, and U_i respects the block structure B_i. Then they define a probability distribution $\Omega = \Omega_{\mathcal{PT}}(\mathcal{V}, \mathcal{B})$ over hidden variable histories $v = (v_0, \cdots, v_T)$ by

$$\Omega_{(v_0,\cdots,v_T)} = \prod_{k=1}^{T} (S_{\mathcal{PT}_{B_k}}(U_{k-1}\cdots U_1|0\rangle^{\otimes\ell}, U_k))_{v_{k-1}v_k}.$$

The oracle $\mathcal{Q}_B$ takes as input the unitaries $U_1, \cdots, U_T$ specified by sequences of gates from some finite universal gate set $\mathcal{U}$. It also takes as input the partitions $B_1, \cdots, B_T$, specified by polynomial-time computable functions $b_1, \cdots, b_T : \{0,1\}^\ell \to \{0,1\}^m$ satisfying the property that x and y are in the same part of the partition B_i if and only if $b_i(x) = b_i(y)$. It outputs a sample from the distribution $\Omega_{\mathcal{PT}}(\mathcal{V}, \mathcal{B})$. Then, let naCQP$'$ be the class of all languages

that can be recognized by a polynomial-time Turing machine with one query to $\mathcal{Q}_B$, with error probability at most $\frac{1}{3}$.

Lemma 7.14.1. $\mathsf{naCQP}' = \mathsf{naCQP}$.

Proof. We first demonstrate a procedure for converting oracle queries to $\mathcal{Q}_B$ to oracle queries to $\mathcal{Q}_P$. Suppose that $B_1, \cdots, B_T$ are specified by polynomial-time computable functions $b_1, \cdots, b_T : \{0,1\}^\ell \to \{0,1\}^m$ (so that x, y are in the same part of the partition B_i if and only if $b_i(x) = b_i(y)$). Now, add an extra T registers of m qubits each, which start in the state $|0 \cdots 0\rangle$. Create a quantum circuit with the same unitary operators $U_1, \cdots, U_T$, but before applying the unitary U_i, apply a unitary that writes the value $|b_i(x)\rangle$ to the ith register when the first ℓ qubits are $|x\rangle$. Then measure the ith register. The effect is that the non-collapsing measurement results will never jump from one part of B_i to a different part, which is exactly what is desired.

To convert a query $C = (U_1, M_1, \cdots, U_T, M_T)$ to $\mathcal{Q}_P$ to a query to $\mathcal{Q}_B$, we first assume, as in the proof of Theorem 7.6.1, that measured qubits are never modified again. Keep the unitaries $U_1, \cdots, U_T$ and let B_i be the partition of $\{0,1\}^\ell$ induced by the measurements $M_1, \cdots, M_{i-1}$. By the principle of deferred measurement, $\Omega_{\mathcal{V},\mathcal{B}}$ is the same distribution that we would have seen had we queried $\mathcal{Q}_P$ instead. □

Now that we have given an alternative definition of naCQP, we can easily show that the choice of gate set does not matter:

Theorem 7.14.1. *Any universal gate set $\mathcal{U}$ yields the same complexity class* naCQP.

Proof. If A is an operator, denote by $\|A\|$ the maximum value of $\|A|\phi\rangle\|_2$ over all ϕ with $\||\phi\rangle\|_2 = 1$.

Lemma 7.14.2. *Suppose that $V_1, \cdots, V_m$ and $V_1', \cdots, V_m'$ are unitary operators. Then,*

$$\left\|V_1 \cdots V_m - V_1' \cdots V_m'\right\| \leq \sum_{k=1}^{m} \left\|V_k - V_k'\right\|.$$

Proof. By induction, it suffices to prove the statement for $m = 2$. We have

$$\begin{aligned}
\left\|V_1V_2 - V_1'V_2'\right\| &= \max_{\||\phi\rangle\|_2=1} \left\|V_1V_2|\phi\rangle - V_1'V_2'|\phi\rangle\right\|_2 \\
&\leq \max_{\||\phi\rangle\|_2=1} \left(\left\|V_1V_2|\phi\rangle - V_1'V_2|\phi\rangle\right\|_2 + \left\|V_1'V_2|\phi\rangle - V_1V_2|\phi\rangle\right\|\right) \\
&= \max_{\||\phi\rangle\|_2=1} \left(\left\|(V_1 - V_1')V_2|\phi\rangle\right\|_2 + \left\|(V_2 - V_2')|\phi\rangle\right\|_2\right) \\
&\leq \left\|V_1 - V_1'\right\| + \left\|V_2 - V_2'\right\|.
\end{aligned}$$

□

If $|\psi\rangle = \Sigma_i \alpha_i |i\rangle$ is a pure state and U is a unitary operator on ℓ qubits that respects the block structure B, such that $U|\psi\rangle = \Sigma_j \beta_j |j\rangle$, then define the joint probabilities matrix $P_{\mathcal{PT}_B}(|\psi\rangle, U)$ by

$$(P_{\mathcal{PT}_B}(|\psi\rangle, U))_{ij} = \begin{cases} \frac{|\alpha_i|^2 |\beta_j|^2}{\sum_{k \sim j} |\beta_k|^2} & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}.$$

It is straightforward to show that

$$\left\| P_{\mathcal{PT}_B}(|\psi\rangle, U) - P_{\mathcal{PT}_B}(|\psi'\rangle, U') \right\|_1 \leq 2^{2\ell}(\left\| |\psi\rangle - |\psi'\rangle \right\|_{tr} + \left\| U - U' \right\|)$$

whenever $|\psi\rangle, |\psi'\rangle$ are state vectors and U, U' are unitary operators.

We use the alternative formulation naCQP' (Lemma 7.14.1). Suppose that $\mathcal{U}$ and $\mathcal{U}'$ are two universal gate sets, and that $\mathcal{V} = (U_1, \cdots, U_T)$ and $\mathcal{B} = (B_1, \cdots, B_T)$ are a query to the $\mathcal{Q}_B$ oracle, where the operators U_t are specified by sequences of gates from $\mathcal{U}$. It is enough to be able to compute in polynomial time a sequence $\mathcal{V}' = (U'_1, \cdots, U'_T)$ of unitary operators, specified by sequences of gates from $\mathcal{U}'$, such that

$$d_{TV}(\Omega_{\mathcal{PT}}(\mathcal{V}, \mathcal{B}), \Omega_{\mathcal{PT}}(\mathcal{V}', \mathcal{B})) < \frac{1}{8}.$$

Let $\varepsilon = 2^{-\ell^2 T - 10}$. Then, by the Solovay-Kitaev theorem [99], it is possible to compute in polynomial time a sequence $\mathcal{V}' = (U'_1, \cdots, U'_T)$ such that

$$\left\| U_t - U'_t \right\| \leq \varepsilon$$

for all t. Suppose that $v = (v_0, \cdots, v_T)$ is sampled from $\Omega_{\mathcal{PT}}(\mathcal{V}, \mathcal{B})$, and that $v' = (v'_0, \cdots, v'_T)$ is sampled from $\Omega_{\mathcal{PT}}(\mathcal{V}', \mathcal{B})$. Then,

$$d_{TV}(\Omega_{\mathcal{PT}}(\mathcal{V}, \mathcal{B}), \Omega_{\mathcal{PT}}(\mathcal{V}', \mathcal{B})) = d_{TV}(v, v').$$

By Lemma 7.6.1,

$$\begin{aligned}
d_{TV}(v, v') &\leq 2 \sum_{i=1}^{T} d_{TV}((v_{i-1}, v_i), (v'_{i-1}, v'_i)) \\
&= 2 \sum_{i=1}^{T} \left\| P_{\mathcal{PT}_{B_i}}(U_{i-1} \cdots U_1 |0\rangle^{\otimes \ell}, U_i) - P_{\mathcal{PT}_{B_i}}(U'_{i-1} \cdots U'_1 |0\rangle^{\otimes \ell}, U'_i) \right\|_1 \\
&\leq 2^{2\ell+1} \sum_{i=1}^{T} \left(\left\| U_{i-1} \cdots U_1 |0\rangle^{\otimes \ell} - U'_{i-1} \cdots U'_1 |0\rangle^{\otimes \ell} \right\|_2 + \left\| U_i - U'_i \right\| \right) \\
&\leq 2^{2\ell+1} \sum_{i=1}^{T} \left(\left\| U_{i-1} \cdots U_1 - U'_{i-1} \cdots U'_1 \right\| + \varepsilon \right) \\
&\leq 2^{2\ell+1} \sum_{i=1}^{T} \left(\sum_{k=0}^{i-1} \left\| U_i - U'_i \right\| + \varepsilon \right) \\
&\leq 2^{2\ell+1} \sum_{i=1}^{T} (T\varepsilon + \varepsilon) \\
&\leq \frac{1}{8},
\end{aligned}$$

as desired. □

7.15 Open questions

We leave many questions about the complexity classes DQP, CQP and naCQP unanswered.

1. We demonstrated a $\tilde{O}(N^{1/3})$-time algorithm for the search problem in the naCQP model, as well as the result that any search algorithm takes $\Omega(N^{1/4})$ time. Is it possible to close the gap between these two bounds? If we disallow intermediate collapsing measurements, then we can prove an $N^{1/3}$ lower bound for search (a proof is included in Section 7.11). However proving an $N^{1/3}$ lower bound when there are intermediate measurements remains open.

2. Can we demonstrate a lower bound, superpolynomial in $\log N$, for the running time of a search algorithm in the DQP model? The proof given in [5] of an $\Omega(N^{1/3})$ lower bound is flawed (as discussed in Section 7.12).

3. Is there a hierarchy of computational models for which the kth allows searching in $\tilde{O}(N^{1/k})$ time?

4. Can we improve the upper bound $\mathsf{naCQP} \subseteq \mathsf{BPP}^{\mathsf{PP}}$ to $\mathsf{naCQP} \subseteq \mathsf{P}^{\mathsf{PP}}$? As we will show in Chapter 8, one cannot place $\mathsf{naCQP} \subseteq \mathsf{PP}$ in a relativizing manner, so we expect it would be difficult to prove $\mathsf{naCQP} \subseteq \mathsf{PP}$. However to our knowledge there is no complexity-theoretic barrier preventing one from placing $\mathsf{naCQP} \subseteq \mathsf{P}^{\mathsf{PP}}$ in a relativizing manner.

5. How powerful is the class CQP? In particular, can one prove a polynomial lower bound for search in CQP? As noted in Appendix 7.8, non-collapsing measurements allow one to (approximately) clone quantum states, which is a non-unitary operation. As we will see shortly in Chapter 9, having the ability to clone *exactly* does allow one to solve NP-hard problems in polynomial time. But it is unclear how to make this algorithm robust to noise, and therefore unclear if NP is contained in CQP.

Chapter 8

On the power of Statistical Zero Knowledge

In the previous chapter, we saw that adding "non-collapsing measurements" to quantum theory allows one to solve problems in SZK in polynomial time in a black-box manner. We then had difficulty proving good upper bounds for this class. In particular, although we could give an upper bound of $\mathsf{BPP}^{\mathsf{PP}}$, we did not manage to give an upper bound of PP, which is the best classical upper bound for BQP. This was because in particular it is open whether or not SZK is contained in PP.

Inspired by this difficulty, in this chapter we turn our attention to studying the structural complexity of zero-knowledge proofs (captured by the complexity class SZK) and their variants. First, we exhibit an oracle relative to which SZK (indeed, even NISZK) is not contained in the class PP. This answers an open question of Watrous from 2002 [1], and in particular implies there can be no relativizing proof of $\mathsf{naCQP} \subseteq \mathsf{PP}$. Second, we give relativized evidence that *perfect* zero knowledge proofs (captured by the class PZK) are weaker than general zero knowledge proofs. Specifically, we exhibit oracles relative to which $\mathsf{SZK} \not\subseteq \mathsf{PZK}$, $\mathsf{NISZK} \not\subseteq \mathsf{NIPZK}$, and $\mathsf{PZK} \neq \mathsf{coPZK}$. The first of these results answers a question raised in 1991 by Aiello and Håstad (Information and Computation), and the second answers a question of Lovett and Zhang (2016). Third, we describe additional applications of these results outside of structural complexity. In particular, we show that this result implies lower bounds for certain parameters of the "polarization lemma" of Sahai and Vadhan [211].

This chapter is based on joint work with Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan [64].

8.1 Introduction

Zero knowledge proof systems, first introduced by Goldwasser, Micali and Rackoff [131], have proven central to the study of complexity theory and cryptography. Abstractly, a zero knowledge proof is a form of interactive proof in which the verifier can efficiently simulate the honest prover on "yes" instances. Therefore, the verifier learns nothing other than whether its input is a "yes" or "no" instance.

In this work, we study *statistical* zero knowledge proofs systems. Here, "efficiently simulate" means that the verifier can, by itself, sample from a distribution which is statistically close to the distribution of the transcript of its interaction with the honest prover[1]. The re-

[1] *Computational* zero-knowledge, in which the zero-knowledge condition is that the verifier can sample

sulting class of decision problems that have statistical zero knowledge proofs is denoted SZK. One can similarly define variants of this class, such as non-interactive statistical zero knowledge (where the proof system is non-interactive, denoted NISZK), or perfect zero knowledge (where the verifier can exactly simulate the honest prover, denoted PZK).

Many problems, some of which are not necessarily in NP, have been shown to admit SZK protocols. These include Graph Non-isomorphism, as well as problems believed to be hard on average, such as Quadratic Residuosity (as well as the closely related discrete logarithm problem), and the Approximate Shortest Vector and Closest Vector problems in lattices [126, 131, 125, 204]. Although SZK contains problems believed to be hard, it lies very low in the polynomial hierarchy (below AM$\cap$coAM), and cannot contain NP-complete problems unless the polynomial hierarchy collapses [118, 34, 62]. Owing in part to its unusual property of containing problems believed to be hard but not NP-complete, SZK has been the subject of intense interest among complexity theorists and cryptographers.

Despite its importance, many basic questions about the hardness of SZK and its variants remain open. Our results in this work can be understood as grouped into three classes, detailed in each of the next three subsections. However, we prove these results via a unified set of techniques.

8.1.1 Group 1: Evidence for the Hardness of SZK

Motivation. Several cryptosystems have been based on the believed hardness of problems in SZK, most notably Quadratic Residuosity and the Approximate Shortest Vector and Closest Vector problems mentioned above. If one could solve SZK-hard problems efficiently, it would break these cryptosystems. Hence, a natural task is to show lower bounds demonstrating that problems in SZK cannot be solved easily. For example, one might want to show that quantum computers or other, more powerful models of computation cannot solve SZK-hard problems efficiently.

Of course, proving such results unconditionally is very difficult, because SZK is contained in AM $\cap$ coAM [118, 34], so even proving lower bounds against classical algorithms solving SZK-hard problems would require separating P from NP.[2] Therefore, a more reasonable goal has been to create oracles relative to which SZK is not contained in other complexity classes; one can then unconditionally prove that "black-box" algorithms from other complexity classes cannot break SZK.

Additional Context. While much progress has been made in this direction (see Section 8.1.5 for details), the problem of giving an oracle separation between SZK and PP has been open since it was posed by Watrous in 2002 [1] and additionally mentioned as an open problem in [8]. Here, PP is the set of decision problems decidable in polynomial time by randomized algorithms with unbounded error. Since a PP algorithm can flip polynomially many coins in its decision process, the gap between the acceptance probabilities of yes and no instances can be exponentially small. PP is a very powerful complexity class – it contains NP and coNP (since it is trivially closed under complement) as well as $\mathsf{BPP}_{\mathsf{path}}$. Furthermore, by Toda's theorem [232], P^{PP} contains the entire polynomial hierarchy. Additionally Aaronson showed PP = PostBQP, the set of problems decidable by quantum algorithms equipped with postselection (the ability to discard all runs of an experiment which do not achieve an

from a distribution that is *computationally indistinguishable* from the transcript, has also been the subject of intense study. In this work we focus exclusively on statistical zero knowledge.

[2] Since SZK $\subseteq$ AM $\cap$ coAM $\subseteq$ PH, if P $\neq$ SZK, then P $\neq$ PH, which in particular implies P $\neq$ NP.

exponentially unlikely outcome). As a result, it is difficult to prove lower bounds against PP.

Our Results. We answer Watrous' question by giving an oracle separating SZK from PP.

Theorem 8.1.1. *There exists an oracle $\mathcal{O}$ such that* $\mathsf{NISZK}^{\mathcal{O}} \not\subset \mathsf{PP}^{\mathcal{O}}$.

In the full version of this work, we actually prove something significantly stronger: our oracle construction separates NISZK from UPP.[3] For the separation from UPP, we refer the reader to the full version of our paper [64].

8.1.2 Group 2: Limitations on the Power of Perfect Zero Knowledge

Motivation. Much progress has been made on understanding the relationship between natural variants of SZK [201, 127, 116, 184, 182]. For example, it is known that SZK = coSZK [201], and if NISZK = coNISZK then SZK = NISZK = coNISZK [127]. Additionally Lovett and Zhang [182] recently gave an oracle separation between NISZK and coNISZK as well as SZK and NISZK. However, many questions remain open, especially regarding the power of *perfect* zero-knowledge proof systems.

Many important SZK protocols, such as the ones for Graph Non-Isomorphism and Quadratic Nonresiduosity, are in fact PZK protocols. This illustrates the power of perfect zero knowledge. In this work, we are primarily concerned with studying the *limitations* of perfect zero knowledge. We are particularly interested in four questions: Does SZK = PZK? What about their non-interactive variants, NISZK and NIPZK? Is PZK closed under complement, the way that SZK is? What about NIPZK? Answering any of these questions in the negative would require showing P ≠ NP, so it is natural to try to exhibit oracles relative to which SZK ≠ PZK, NISZK ≠ NIPZK, PZK ≠ coPZK, and NIPZK ≠ coNIPZK.

Additional Context. In 1991, Aiello and Håstad [33] gave evidence that PZK contains hard problems by creating an oracle relative to which PZK is not contained in BPP. On the other hand, they also gave an oracle that they *conjectured* separates SZK from PZK (but were unable to prove this). Exhibiting such an oracle requires a technique that can tell the difference between zero simulation error (PZK) and simulation to inverse exponential error (SZK), and prior to our work, no such technique was known. The question of whether SZK = PZK has been asked by Goldwasser [130] as well. The analogous question for the non-interactive classes NISZK and NIPZK is also well motivated, and was explicitly asked in recent work of Lovett and Zhang [182].

Determining whether variants of SZK satisfy the same closure properties as SZK is natural as well: indeed, a main result of Lovett and Zhang [182] is an oracle relative to which NISZK ≠ coNISZK.

Our Results. We give oracles separating SZK from PZK, NISZK from NIPZK, PZK from coPZK, and NIPZK from coNIPZK. The first two results answer the aforementioned questions raised by Aiello and Håstad [33] (though our oracle is different from the candidate proposed by Aiello and Håstad), and Lovett and Zhang [182]. Along the way, we show that PZK is contained in PP in a relativizing manner – this is in sharp contrast to SZK (see Theorem 8.1.1).

[3] UPP is traditionally defined as an oracle complexity class, in which machines must output the correct answer with probability strictly greater than 1/2, and are charged for oracle queries but not for computation time. In this model, the gap between 1/2 and the probability of outputting the correct answer can be *arbitrarily* (in particular, superexponentially) small.

Theorem 8.1.2. *For any oracle $\mathcal{O}$, $\mathsf{PZK}^{\mathcal{O}} \subseteq \mathsf{PP}^{\mathcal{O}}$. In addition, there exist oracles $\mathcal{O}_1$ and $\mathcal{O}_2$ such that $\mathsf{SZK}^{\mathcal{O}_1} \not\subseteq \mathsf{PZK}^{\mathcal{O}_1}$, $\mathsf{NISZK}^{\mathcal{O}_1} \not\subseteq \mathsf{NIPZK}^{\mathcal{O}_1}$, $\mathsf{PZK}^{\mathcal{O}_2} \not\subseteq \mathsf{coPZK}^{\mathcal{O}_2}$, and $\mathsf{NIPZK}^{\mathcal{O}_2} \not\subseteq \mathsf{coNIPZK}^{\mathcal{O}_2}$.*

A summary of known relationships between complexity classes in the vicinity of SZK, including the new results established in this work, is provided in Figure 8-1.

8.1.3 Group 3: Consequences for Polarization and Property Testing

In addition to the above oracle separations, our results have a number of applications in other areas of theoretical computer science. For example, as previously mentioned, our results have implications regarding the power of complexity classes capturing the power of quantum computing with "more powerful" modified versions of quantum mechanics [13, 5]. In particular they imply that allowing for "non-collapsing measurements" may be more powerful than allowing for postselection in quantum computing, as the former ability allows one to solve SZK-hard problems, and the latter modification is contained in PP. Therefore this modification seems to be quite powerful.

We also show that these results imply limitations on the Polarization Lemma of Sahai and Vadhan [211]. The Polarization Lemma shows how to amplify differences in statistical distance. In particular, suppose one has two distributions D_0 and D_1 such that either $||D_0 - D_1|| \leq 1/3$ or if $||D_0 - D_1|| \geq 2/3$, where $||D_0 - D_1||$ denotes total variation distance. Then, in polynomial time and using only black-box samples from D_0 and D_1, one can produce distributions D_0' and D_1' such that, if $||D_0 - D_1|| \leq 1/3$, then $||D_0' - D_1'|| < \varepsilon$, and if $||D_0 - D_1|| \geq 2/3$, then $||D_0' - D_1'|| > 1 - \varepsilon$, where ε is exponentially small. In other words their transformation "polarizes" the distributions to be either very close or very far from one another. This is a key part of their proof that the Statistical Difference problem is complete for SZK.

As one pushes ε to smaller values in the Polarization Lemma, the size of the range of the output distributions D_0', D_1' increases. Our oracle separation of SZK from PP implies lower bounds on the size of ε relative to the size of the output distributions n'. We obtain these by showing that if ε can be made very small relative to n', then that would place $\mathsf{SZK}^{\mathcal{O}} \subseteq \mathsf{PP}^{\mathcal{O}}$ (and even $\mathsf{SZK}^{\mathcal{O}} \subseteq \mathsf{BPP}_{\mathsf{path}}{}^{\mathcal{O}}$) for all oracles $\mathcal{O}$. Therefore, as a corollary of our main result, ε cannot be made very small by any poly-time black-box polarization algorithm. More specifically, if n' is the number of bits in the output range of D_0', D_1', then we achieve a lower bound of $\varepsilon > 2^{-n'/2-1}$ for any poly-time polarization algorithm:

Theorem 8.1.3. *There does not exist an algorithm running in poly(n) time, which given black box distributions D_0, D_1 on strings of length n which obey either $||D_0 - D_1|| < 1/3$ or $||D_0 - D_1|| > 2/3$, produces two output distributions D_0' and D_1' on strings of length $n' = poly(n)$ such that either $||D_0' - D_1'|| < \varepsilon$ (in the first case) or $||D_0' - D_1'|| > 1 - \varepsilon$ (in the second case) where $\varepsilon \leq 2^{-n'/2-1}$.*

We provide a proof of this theorem in Section 8.6. Note that our lower bounds apply to *any* form of poly-time black box polarization. To our knowledge, the only other known lower bound against arbitrary black-box zero-knowledge reductions is in concurrent results of Lovett and Zhang [182] regarding the impossibility of reversing entropy approximation. Prior to these works, lower bounds were only known against particular forms of polarization. For instance, Holenstein and Renner proved that if one is taking independent samples from

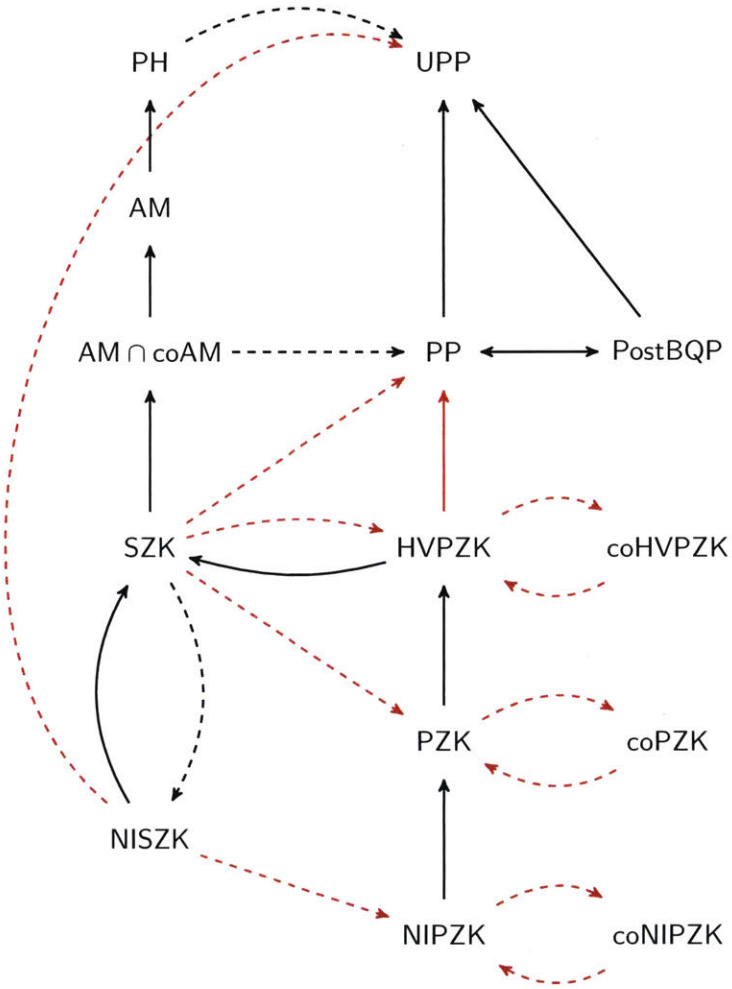


Figure 8-1: $\mathcal{C}_1 \to \mathcal{C}_2$ indicates $\mathcal{C}_1$ is contained in $\mathcal{C}_2$ respect to *every* oracle, and $\mathcal{C}_1 \dashrightarrow \mathcal{C}_2$ denotes that there is an oracle $\mathcal{O}$ such that $\mathcal{C}_1^{\mathcal{O}} \not\subset \mathcal{C}_2^{\mathcal{O}}$. Red indicates new results. Certain non-inclusions that are depicted are subsumed by other non-inclusions (e.g., NISZK not in UPP subsumes SZK not in PP). We include some redundant arrows to facilitate comparison of our results to prior work.

the input distributions[4], then one cannot polarize distributions which are $\geq \alpha$ far or $\leq \beta$ close unless $\alpha^2 > \beta$ [154]. Note that if one wishes to prove our lower bounds for Holenstein-Renner style Polarization only, then there is a more direct proof of this fact using Fourier analysis. We provide this simplified proof in Section 8.6.3 to facilitate understanding of why this sort of polarization is impossible. We also note that it remains open to close the gap between our lower bound of $\varepsilon = 2^{-n'/2-1}$ and the upper bound given by Sahai and Vadhan, which is $\varepsilon = 2^{-n'^{1/2+\delta}}$ for any $\delta > 0$ [211].

Property Testing: Additionally, our results yield novel lower bounds for certain forms of property testing algorithms. We will describe this in detail in Section 8.6 and Section 8.7. Our results also imply upper bounds for *streaming interactive proofs* [94, 86]. We refer the reader to the full version of our work [64] for details of the latter consequence.

8.1.4 Overview of Our Techniques

Oracle Separation of NISZK and PP (Proof Overview for Theorem 8.1.1)

To describe our methods, it is helpful to introduce the notions of approximate degree and threshold degree, both of which are measures of Boolean function complexity that capture the difficulty of point-wise approximation by low-degree polynomials. The ε-approximate degree of a function $f\colon \{0,1\}^n \to \{0,1\}$, denoted $\widetilde{\deg}_\varepsilon(f)$, is the least degree of a real polynomial that point-wise approximates f to error ε. The threshold degree of f, denoted $\deg_\pm(f)$, is the least degree of a real polynomial that agrees in sign with f at all points. It is easy to see that threshold degree is equivalent to the limit of the approximate degree as the error parameter ε approaches $1/2$ from below.

A recent and growing line of work has addressed a variety of open problems in complexity theory by establishing various forms of hardness amplification for approximate degree. Roughly speaking, these results show how to take a function f which is hard to approximate by degree d polynomials to error $\varepsilon = 1/3$, and turn f into a related function F that is hard to approximate by degree d polynomials even when ε is very close to $1/2$. In most of these works, F is obtained from f by block-composing f with a "hardness-amplifying function" g. We denote such a block-composition by $g(f)$.

The technical core of our result lies in establishing a new form of hardness amplification for approximate degree. Specifically, let g be the partial function $\mathsf{GapMaj}_n\colon \{0,1\}^n \to \{0,1\}$ (throughout this introduction, whenever necessary, we use subscripts after function names to clarify the number of variables on which the function is defined). Here GapMaj is the gapped majority function, defined, for some $1 \geq \delta > 0.5$, to be 1 if $\geq \delta$ fraction of its inputs are 1, to be 0 if $\geq \delta$ fraction of its inputs are 0, and to be undefined otherwise (in this introduction, we will ignore the precise choice of δ that we use in our formal results).[5]

Theorem 8.1.4. *(Informal) Let* $f\colon \{0,1\}^M \to \{0,1\}$. *Suppose that* $\widetilde{\deg}_{1/3}(f) \geq d$. *Define* $F\colon \{0,1\}^{n \cdot M} \to \{0,1\}$ *via* $F = \mathsf{GapMaj}_n(f)$. *Then* $\deg_\pm(F) = \Omega(\min(d, n))$.

In our main application of Theorem 8.1.4, we apply the theorem to a well-known (partial) function $f = \mathsf{Col}_M$ called the Collision problem. This function is known to have approximate

[4]Their result also allows one to append a random string r, drawn without querying the input distributions D_0 or D_1.

[5]We clarify that if f is a partial function then $\mathsf{GapMaj}_n(f)$ is technically not a composition of functions, since for some inputs $x = (x_1, \ldots, x_n)$ on which $\mathsf{GapMaj}_n(f)$ is defined, there may be values of i for which x_i is outside of the domain of f. See Section 8.2.4 for further discussion of this point.

degree $\tilde{\Omega}(M^{1/3})$, so Theorem 8.1.4 implies that $F := \mathsf{GapMaj}_{M^{1/3}}(\mathsf{Col}_M)$ has threshold degree $\tilde{\Omega}(M^{1/3})$. Standard results then imply that the UPP *query complexity* of F is $\tilde{\Omega}(M^{1/3})$ as well. That is, $F \notin \mathsf{UPP}^{\mathsf{dt}}$.

Corollary 8.1.5 (Informal). *Let $m = M^{4/3}$, and define $F\colon \{0,1\}^m \to \{0,1\}$ via $F := \mathsf{GapMaj}_{M^{1/3}}(\mathsf{Col}_M)$. Then $\mathsf{UPP}^{\mathsf{dt}}(F) = \tilde{\Omega}(m^{1/4})$.*

Moreover, as we show later, $\mathsf{GapMaj}_{M^{1/3}}(\mathsf{Col}_M)$ is in $\mathsf{NISZK}^{\mathsf{dt}}$. Hence, we obtain a separation between $\mathsf{NISZK}^{\mathsf{dt}}$ and $\mathsf{UPP}^{\mathsf{dt}}$. The desired oracle separating NISZK from UPP follows via standard methods.

In this thesis, we will merely prove a slightly simpler result - namely a composition theorem lifting approximate degree to discrepancy. Discrepancy is a measure which captures the complexity class PP, in the same way that threshold degree captures the complexity class UPP. We refer readers interested in the UPP result to refer to the full version of our paper [64].

Comparison of Theorem 8.1.4 to Prior Work. The hardness amplification result from prior work that is most closely related to Theorem 8.1.4 is due to Sherstov [218]. Sherstov's result makes use of a notion known as (positive) one-sided approximate degree [218, 81]. Positive one-sided approximate degree is a measure that is intermediate between approximate degree and threshold degree – the positive one-sided approximate degree of f, denoted $\deg^+_\varepsilon(f)$, is always at most as large as the approximate degree of f but can be much smaller, and it is always at least as large as the threshold degree of f but can be much larger (see Section 8.2.2 for a formal definition of positive one-sided approximate degree).[6]

Theorem 8.1.6 (Sherstov). *Let $f\colon \{0,1\}^M \to \{0,1\}$. Suppose that $\deg^+_{1/3}(f) \geq d$. Define $F\colon \{0,1\}^{n \cdot M} \to \{0,1\}$ via $F = \mathsf{AND}_n(f)$. Then $\deg_\pm(F) = \Omega(\min(d,n))$.*[7]

There are two differences between Theorems 8.1.4 and 8.1.6. The first is that the hardness-amplifier in Theorem 8.1.4 is GapMaj, while in Theorem 8.1.6 it is AND. GapMaj is a "simpler" function than AND in the following sense: block-composing f with GapMaj preserves membership in complexity classes such as $\mathsf{NISZK}^{\mathsf{dt}}$ and $\mathsf{SZK}^{\mathsf{dt}}$; this is not the case for AND, as AND itself is not in $\mathsf{SZK}^{\mathsf{dt}}$. This property is essential for us to obtain threshold degree lower bounds even for functions that are in $\mathsf{NISZK}^{\mathsf{dt}}$.

The second difference is that Theorem 8.1.4 holds under the assumption that $\widetilde{\deg}_{1/3}(f) \geq d$, while Theorem 8.1.6 makes the stronger assumption that $\deg^+_\varepsilon(f) \geq d$. While we do not exploit this second difference in our applications, ours is the first form of hardness amplification that works for approximate degree rather than one-sided approximate degree. This property has already been exploited in subsequent work [82].

Proof Sketch for Theorem 8.1.4. A *dual polynomial* is a dual solution to an appropriate linear program capturing the threshold degree of any function. Specifically, for a

[6] The notion of positive one-sided approximate degree treats inputs in $f^{-1}(1)$ and $f^{-1}(0)$ asymmetrically. There is an analogous notion called negative one-sided approximate degree that reverses the roles of $f^{-1}(1)$ and $f^{-1}(0)$ [231, 164]. Our use of the positive vs. negative terminology follows prior work [231, 164] - other prior works [218, 81] only used negative one-sided approximate degree, and referred to this complexity measure without qualification as one-sided approximate degree. In this thesis, we exclusively use the notion of positive one-sided approximate degree.

[7] Sherstov stated his result for $\mathsf{OR}_n(f)$ under the assumption that f has large *negative* one-sided approximate degree. Our statement of Theorem 8.1.6 is the equivalent result under the assumption that f has large positive one-sided approximate degree.

(partial) function f defined on a subset of $\{0,1\}^n$, a dual polynomial witnessing the fact that $\widetilde{\deg}_\varepsilon(f) \geq d$ is a function $\psi\colon \{0,1\}^n \to \mathbb{R}$ that satisfies the following three properties.

(a) ψ is uncorrelated with all polynomials p of total degree at most d. That is, for any $p\colon \{0,1\}^n \to \mathbb{R}$ such that $\deg(p) \leq d$, it holds that $\sum_{x\in\{0,1\}^n} \psi(x) \cdot p(x) = 0$. We refer to this property by saying that ψ has *pure high degree* d.

(b) ψ has ℓ_1 norm equal to 1, i.e., $\sum_{x\in\{0,1\}^n} |\psi(x)| = 1$.

(c) ψ has correlation at least ε with f. That is, if D denotes the domain on which f is defined, then $\sum_{x\in D} \psi(x) \cdot f(x) - \sum_{x\in\{0,1\}^n\setminus D} |\psi(x)| > \varepsilon$.

It is not hard to see that a dual witness for the fact that $\deg_\pm(f) \geq d$ is a function ψ satisfying Properties (a) and (b) above, that additionally is *perfectly* correlated with f. That is, ψ additionally satisfies

$$\sum_{x\in D} \psi(x) \cdot f(x) - \sum_{x\in\{0,1\}^n\setminus D} |\psi(x)| = 1. \tag{8.1}$$

In this case, $\psi \cdot f$ is non-negative, and is referred to as an *orthogonalizing distribution* for f.

We prove Theorem 8.1.4 by constructing an explicit orthogonalizing distribution for $\mathsf{GapMaj}_n(f)$. Specifically, we show how to take a dual polynomial witnessing the fact that $\widetilde{\deg}_{1/3}(f) \geq d$, and turn it into an orthogonalizing distribution witnessing the fact that $\deg_\pm(F) = \Omega(\min(d,n))$. Our construction of an orthogonalizing distribution for $\mathsf{GapMaj}_n(f)$ is inspired by and reminiscent of Sherstov's construction of an orthogonalizing distribution for $\mathsf{AND}_n(f)$ [218], which in turn builds on a dual polynomial for $\mathsf{AND}_n(f)$ constructed by Bun and Thaler [81].

Limitations on the Power of Perfect Zero Knowledge (Proof Overview For Theorem 8.1.2)

We begin the proof of Theorem 8.1.2 by showing that HVPZK (*honest verifier* perfect zero knowledge) is contained in PP in a relativizing manner (see Section 8.5). Since the inclusions PP $\subseteq$ UPP, NIPZK $\subseteq$ HVPZK, PZK $\subseteq$ HVPZK, and NISZK $\subseteq$ SZK hold with respect to any oracle, this means that our oracle separating NISZK from UPP (Theorem 8.1.1) also separates SZK from PZK and NISZK from NIPZK.

We then turn to showing that PZK and NIPZK are not closed under complement with respect to some oracle. Since the proofs are similar, we focus on the case of PZK in this overview.

Since both PZK and coPZK are contained in PP with respect to any oracle, our oracle separation of NISZK from PP (Theorem 8.1.1) does not imply an oracle relative to which PZK $\neq$ coPZK. Instead, to obtain this result we prove a new amplification theorem for one-sided approximate degree. Using similar techniques as Theorem 8.1.4, we show that if f has high positive one-sided approximate degree, then block-composing f with the gapped AND function yields a function with high threshold degree. Here GapAND is the partial function that outputs 1 if all inputs are 1, outputs 0 if at least a δ fraction of inputs are 0, and is undefined otherwise.

Theorem 8.1.7. *(Informal) Let* $f\colon \{0,1\}^M \to \{0,1\}$. *Suppose that* $\deg^+_{1/3}(f) \geq d$. *Then* $\deg_\pm(\mathsf{GapAND}_n(f)) = \Omega(\min(d,n))$.

We then show that (a) $\mathsf{PZK}^{\mathsf{dt}}$ is closed under composition with GapAND and (b) there is a function f in $\mathsf{PZK}^{\mathsf{dt}}$ whose *complement* $\bar{f}$ has high positive one-sided approximate degree. If $\mathsf{PZK}^{\mathsf{dt}}$ were closed under complement, then $\bar{f}$ would be in $\mathsf{PZK}^{\mathsf{dt}}$. By amplifying the hardness of $\bar{f}$ using Theorem 8.1.7, we obtain a problem that is still in $\mathsf{PZK}^{\mathsf{dt}}$ (this holds by property (a)) yet outside of $\mathsf{PP}^{\mathsf{dt}}$ (this holds by property (b), together with Theorem 8.1.7). This is easily seen to contradict the fact PZK is in PP relative to all oracles. Hence, $\bar{f}$ is a function in $\mathsf{coPZK}^{\mathsf{dt}}$ that is not in $\mathsf{PZK}^{\mathsf{dt}}$, and standard techniques translate this fact into an oracle separating coPZK from PZK. We provide details of these results in Section 8.5.

8.1.5 Other Works Giving Evidence for the Hardness of SZK

As mentioned in Section 8.1.2, Aiello and Håstad showed that PZK (and also SZK) is not contained in BPP relative to some oracle [33]. Agrawal et al. later used similar techniques to show that SZK is not contained in the class SRE (which can be viewed as a natural generalization of BPP) relative to some oracle [26]. Aaronson [2] gave an oracle relative to which SZK is not contained in BQP – and therefore quantum computers cannot break SZK-hard cryptosystems in a black-box manner. Building on that work, Aaronson [8] later gave oracle separations against the class QMA (a quantum analogue of NP) and the class $\mathsf{A_0PP}$ (a class intermediate between QMA and PP). Therefore even quantum proofs cannot certify SZK in a black-box manner[8].

Until recently, the lower bound most closely related to our oracle separation of NISZK and UPP (cf. Theorem 8.1.1) was Vereshchagin's result from 1995, which gave an oracle relative to which $\mathsf{AM} \cap \mathsf{coAM}$ is not contained in PP [238]. Our result is an improvement on Vereshchagin's because the inclusions $\mathsf{NISZK} \subseteq \mathsf{SZK} \subseteq \mathsf{AM} \cap \mathsf{coAM}$ can be proved in a relativizing manner (cf. Figure 8-1). It also generalizes Aaronson's oracle separation between SZK and $\mathsf{A_0PP}$ [8].

Vereshchagin [238] also reports that Beigel claimed a simple proof of the existence of a function f that is in the query complexity class $\mathsf{AM}^{\mathsf{dt}}$, but is not in the query complexity class $\mathsf{UPP}^{\mathsf{dt}}$. Our result improves on Beigel's in two regards. First, since $\mathsf{NISZK}^{\mathsf{dt}} \subseteq \mathsf{AM}^{\mathsf{dt}}$, separating $\mathsf{NISZK}^{\mathsf{dt}}$ from $\mathsf{UPP}^{\mathsf{dt}}$ is more difficult than separating $\mathsf{AM}^{\mathsf{dt}}$ from $\mathsf{UPP}^{\mathsf{dt}}$. Second, Beigel only claimed a superlogarithmic lower bound on the $\mathsf{UPP}^{\mathsf{dt}}$ query complexity of f, while we give a polynomial lower bound.

Theorem 8.1.1 also improves on very recent work of Chen [88, 89], which gave a query separation between the classes $\mathsf{P}^{\mathsf{SZK}}$ and PP.

Finally note that under a strong derandomization hypothesis, AM = NP [192], which would place SZK ⊆ PP. It is unclear if this derandomization hypothesis is true, or if SZK ⊆ PP in the unrelativized world.

[8] Note, however, that oracle separations do not necessarily imply the analogous separations in the "real world" – see [48] and [87] for instances in which the situation in the presence of oracles is far from the situation in the real world.

8.2 Technical Preliminaries

8.2.1 Complexity Classes

Notation. In an interactive proof (P, V) where P is the prover and V is the verifier, we denote by $(P, V)(x)$ the random variable corresponding to the transcript of the protocol on input x. For distributions D_0 and D_1, $||D_0 - D_1||$ denotes the Total Variational Distance between them $\left(||D_0 - D_1|| = \frac{1}{2}|D_0 - D_1|_1\right)$.

Definition 8.2.1 (Honest Verifier Statistical Zero Knowledge)**.** A promise problem $L = (L_Y, L_N)$ is in HVSZK if there exists a tuple of Turing machines (P, V, S), where the *verifier* V and *simulator* S run in probabilistic polynomial time, satisfying the following:

- (P, V) is an interactive proof for L with negligible completeness and soundness errors.
- For any $x \in L_Y$,

$$||S(x) - (P, V)(x)|| \leq \mathsf{negl}(|x|)$$

Definition 8.2.2 (Non-Interactive SZK)**.** A promise problem $L = (L_Y, L_N)$ is in NISZK if there exist a tuple of Turing machines (P, V, S), where the *verifier* V and *simulator* S run in probabilistic polynomial time, satisfying the following:

- (P, V) is an interactive proof for L with negligible completeness and soundness errors, with the following additional conditions:
  1. P and V both have access to a long enough common random string.
  2. The interactive proof consists of a single message from P to V.
- For any $x \in L_Y$,

$$||S(x) - (P, V)(x)|| \leq \mathsf{negl}(|x|)$$

The definitions of these classes in the presence of an oracle are the same, except that P, V, and S all have access to the oracle.

It is easy to see that NISZK is contained in HVSZK, even in the presence of oracles. The class SZK is defined to be almost the same as HVSZK, except for the stipulation that for any verifier (even one that deviates from the prescribed protocol), there is a simulator that simulates the prover's interaction with that verifier. It was shown in [128] that SZK is equal to HVSZK, and their proof continues to hold in the presence of any oracle. For this reason, NISZK is also contained in SZK in the presence of any oracle, and we shall be implicitly making use of this fact at several points where we state corollaries for SZK instead of HVSZK.

8.2.2 Approximate Degree, Threshold Degree, and Their Dual Characterizations

We first recall the definitions of approximate degree, positive one-sided approximate degree, and threshold degree for partial functions.

Definition 8.2.3. Let $D \subseteq \{0,1\}^M$, and let f be a function mapping D to $\{0,1\}$.

- The *approximate degree* of f with approximation constant $0 \leq \varepsilon < 1/2$, denoted $\widetilde{\deg}_\varepsilon(f)$, is the least degree of a real polynomial $p\colon \{0,1\}^M \to \mathbb{R}$ such that $|p(x) - f(x)| \leq \varepsilon$ when $x \in D$, and $|p(x)| \leq 1 + \varepsilon$ for all $x \notin D$. We refer to such a p as an *approximating polynomial* for f. We use $\widetilde{\deg}(f)$ to denote $\widetilde{\deg}_{1/3}(f)$.
- The *threshold degree* of f, denoted $\deg_\pm(f)$, is the least degree of a real polynomial p such that $p(x) > 0$ when $f(x) = 1$, and $p(x) < 0$ when $f(x) = 0$.
- The *postive one-sided approximate degree* of f with approximation constant $0 \leq \varepsilon < 1/2$, denoted $\deg^+_\varepsilon(f)$, is the least degree of a real polynomial p such that $|p(x) - 1| \leq \varepsilon$ for all $x \in f^{-1}(1)$, and $p(x) \leq \varepsilon$ when $x \in f^{-1}(0)$. We refer to such a p as a *positive one-sided approximating polynomial* for f. We use $\deg^+(f)$ to denote $\deg^+_{1/3}(f)$.

Remark. We highlight the following subtlety in Definition 8.2.3: an approximating polynomial for a partial function f is required to be bounded in absolute value even outside of the domain D on which f is defined, yet this is not required of a one-sided approximating polynomial for f. The reason we choose to require an approximating polynomial to be bounded outside of D is to ensure that the Col function (defined later in Section 8.2.5) has large approximate degree.

There are clean dual characterizations for each of the three quantities defined in Definition 8.2.3. We state these characterizations without proof, and direct the interested reader to [219, 218, 80] for details.

For a function $\psi\colon \{0,1\}^M \to \mathbb{R}$, define the ℓ_1 norm of ψ by $\|\psi\|_1 = \sum_{x \in \{0,1\}^M} |\psi(x)|$. If the support of a function $\psi\colon \{0,1\}^M \to \mathbb{R}$ is (a subset of) a set $D \subseteq \{0,1\}^M$, we will write $\psi\colon D \to \mathbb{R}$. For functions $f, \psi\colon D \to \mathbb{R}$, denote their inner product by $\langle f, \psi \rangle := \sum_{x \in D} f(x)\psi(x)$. We say that a function $\psi\colon \{0,1\}^M \to \mathbb{R}$ has *pure high degree* d if ψ is uncorrelated with any polynomial $p\colon \{0,1\}^M \to \mathbb{R}$ of total degree at most d, i.e., if $\langle \psi, p \rangle = 0$.

Theorem 8.2.4. *Let $f : D \to \{0,1\}$ with $D \subseteq \{0,1\}^M$ be a partial function and ε be a real number in $[0, 1/2)$. $\widetilde{\deg}_\varepsilon(f) > d$ if and only if there is a real function $\psi : \{0,1\}^M \to \mathbb{R}$ such that:*

1. *(Pure high degree): ψ has pure high degree of d.*
2. *(Unit ℓ_1-norm): $\|\psi\|_1 = 1$.*
3. *(Correlation): $\sum_{x \in D} \psi(x) f(x) - \sum_{x \notin D} |\psi(x)| > \varepsilon$.*

Theorem 8.2.5. *Let $f : D \to \{0,1\}$ with $D \subseteq \{0,1\}^M$ be a partial function. $\deg_\pm(f) > d$ if and only if there is a real function $\psi : D \to \mathbb{R}$ such that:*

1. *(Pure high degree): ψ has pure high degree of d.*
2. *(Sign Agreement): $\psi(x) \geq 0$ when $f(x) = 1$, and $\psi(x) \leq 0$ when $f(x) = 0$.*
3. *(Non-triviality): $\|\psi\|_1 > 0$.*

Theorem 8.2.6. *Let $f : D \to \{0,1\}$ with $D \subseteq \{0,1\}^M$ be a partial function and ε be a constant in $[0,1/2)$. $\deg^+_\varepsilon(f) > d$ if and only if there is a real function $\psi : D \to \mathbb{R}$ such that:*

1. *(Pure high degree): ψ has pure high degree of d.*
2. *(Unit ℓ_1-norm): $\|\psi\|_1 = 1$.*
3. *(Correlation): $\langle \psi, f \rangle > \varepsilon$.*
4. *(Negative Sign Agreement): $\psi(x) \leq 0$ whenever $f(x) = 0$.*

8.2.3 $\mathsf{PP}^{\mathsf{dt}}$ and $\mathsf{UPP}^{\mathsf{dt}}$

Now we define the two natural analogues of PP complexity in the query model.

Definition 8.2.7. Let $f : D \to \{0,1\}$ with $D \subseteq \{0,1\}^M$ be a partial function. Let $\mathcal{T}$ be a randomized decision tree which computes f with a probability better than $1/2$. Let α be the maximum real number such that

$$\min_{x \in D} \Pr[\mathcal{T} \text{ outputs } f(x) \text{ on input } x] \geq \frac{1}{2} + \alpha.$$

Then we define the PP query cost of $\mathcal{T}$ for f to be $\mathsf{PP}^{\mathsf{dt}}(\mathcal{T}; f) = C(\mathcal{T}; f) + \log_2(1/\alpha)$, where $C(\mathcal{T}; f)$ denotes the maximum number of queries $\mathcal{T}$ incurs on an input in the worst case. We define $\mathsf{UPP}^{\mathsf{dt}}(\mathcal{T}; f) = C(\mathcal{T}; f)$. Observe that $\mathsf{UPP}^{\mathsf{dt}}(\mathcal{T}; f)$ is the same as $\mathsf{PP}^{\mathsf{dt}}(\mathcal{T}; f)$, except that the advantage α of the randomized decision tree over random guessing is not incorporated into $\mathsf{UPP}^{\mathsf{dt}}(\mathcal{T}; f)$. We define $\mathsf{PP}^{\mathsf{dt}}(f)$ (respectively, $\mathsf{UPP}^{\mathsf{dt}}$) as the minimum of $\mathsf{PP}^{\mathsf{dt}}(\mathcal{T}; f)$ (respectively, $\mathsf{UPP}^{\mathsf{dt}}(\mathcal{T}; f)$) over all $\mathcal{T}$ that computes f with a probability better than $1/2$.

$\mathsf{PP}^{\mathsf{dt}}$ is closely related to approximate degree with error very close to $1/2$. We have the following well-known relationship between them.

Lemma 8.2.8. *Let $f : D \to \{0,1\}$ with $D \subseteq \{0,1\}^M$ be a partial function. Suppose $\widetilde{\deg}_{1/2-2^{-d}}(f) > d$ for some positive integer d. Then $\mathsf{PP}^{\mathsf{dt}}(f) > d/2$.*

Meanwhile, $\mathsf{UPP}^{\mathsf{dt}}$ is exactly characterized by threshold degree.

Lemma 8.2.9. *Let $f : D \to \{0,1\}$ with $D \subseteq \{0,1\}^M$ be a partial function. Then $\mathsf{UPP}^{\mathsf{dt}}(f) = \deg_\pm(f)$.*

8.2.4 Gap Majority and Gap AND

In this subsection we introduce a transformation of partial functions which will be used in this paper.

Definition 8.2.10. Let $f : D \to \{0,1\}$ with $D \subseteq \{0,1\}^M$ be a partial function and n be a positive integer, $0.5 < \varepsilon \leq 1$ be a real number. We define the gap majority version of f, denoted by $\mathsf{GapMaj}_{n,\varepsilon}(f)$, as follows:

Given an input $x = (x_1, x_2, \ldots, x_n) \in \{0,1\}^{M \cdot n}$, we define $n_{\mathsf{Yes}}(x) := \sum_{i=1}^{n} \mathbb{1}_{x_i \in D \wedge f(x_i)=1}$ and

$n_{\mathsf{No}}(x) := \sum_{i=1}^{n} \mathbb{1}_{x_i \in D \wedge f(x_i)=0}$. Then

$$\mathsf{GapMaj}_{n,\varepsilon}(f)(x) = \begin{cases} 1 & \text{when } n_{\mathsf{Yes}}(x) \geq \varepsilon \cdot n \\ 0 & \text{when } n_{\mathsf{No}}(x) \geq \varepsilon \cdot n \\ \text{undefined} & \text{otherwise} \end{cases}$$

Note that even on inputs x for which $\mathsf{GapMaj}_{n,\varepsilon}(f)(x)$ is defined, there may be some values of i for which x_i is not in D. For brevity, we will occasionally write $\mathsf{GapMaj}(f)$ when n and ε are clear from context.

We also define the GapAND function. This is a partial function that agrees with the total function AND wherever it is defined.

Definition 8.2.11. Let n be a positive integer, $0 < \varepsilon < 1$ be a constant. We define the Gapped AND function, $\mathsf{GapAND}_{n,\varepsilon} : D \to \{0,1\}$ with $D \subseteq \{0,1\}^n$, as the function that outputs 1 if all inputs are 1; outputs 0 if at least $\varepsilon \cdot n$ inputs are 0; and is undefined otherwise.

For a partial function $f : D \to \{0,1\}$ with $D \subseteq \{0,1\}^M$, we define $\mathsf{GapAND}_{n,\varepsilon}(f)$ to be a true block-composition of partial functions, i.e.,

$$\mathsf{GapAND}_{n,\varepsilon}(f)(x_1, \ldots, x_n) = \mathsf{GapAND}_{n,\varepsilon}(f(x_1), \ldots, f(x_n))$$

whenever the right hand side of the equality is defined, and $\mathsf{GapAND}_{n,\varepsilon}(f)$ is undefined otherwise.

Remark 8.2.12. *Note that* $\mathsf{GapMaj}_{n,\varepsilon}(f)$ *is not technically a block-composition of partial functions, since* $\mathsf{GapMaj}_{n,\varepsilon}(f)(x_1, \ldots, x_n)$ *is defined even on some inputs for which some* $f(x_i)$ *is not defined.*

8.2.5 Problems

We now recall the Collision problem. This problem interprets its input as a function f mapping $[n]$ to $[n]$, and the goal is to decide whether the input is a permutation or is 2-to-1, promised that one of them is the case. We need a slightly generalized version, which asks to distinguish between permutations and k-to-1 functions.

Definition 8.2.13 (Collision problem)**.** Fix an integer $k \geq 2$, and assume for simplicity that n is a power of 2. The partial function Col_n^k is defined on a subset of $\{0,1\}^{n \log n}$. It interprets its input as specifying a function $f\colon [n] \to [n]$ in the natural way, and evaluates to 1 if f is a permutation, 0 if f is a k-to-1 function, and is undefined otherwise. When k and n are clear from context, we write Col for brevity.

This problem admits a simple SZK protocol in which the verifier makes only polylog(n) queries to the input. Specifically, the verifier executes the following sub-protocol polylog(n) times: the verifier chooses a random $i \in [n]$, makes a single query to learn $f(i)$, sends $f(i)$ to the prover, and rejects if the prover fails to respond with i. It is easy to see that the sub-protocol has perfect completeness, constant soundness error, and is perfect zero knowledge. Because the sub-protocol is repeated polylog(n) times, the total soundness error is negligible.

In 2002, Aaronson [2] proved the first non-constant lower bound for the Col_n^2 problem: namely, any bounded-error quantum algorithm to solve it needs $\Omega(n^{1/5})$ queries to f. Aaronson and Shi [19] subsequently improved the lower bound to $\Omega(n^{1/3})$, for functions $f : [n] \to [3n/2]$; then Ambainis [36] and Kutin [177] proved the optimal $\Omega(n^{1/3})$ lower bound for functions $f : [n] \to [n]$.

We need a version of the lower bound that makes explicit the dependence on k and ε.

Theorem 8.2.14 (Implicit in Kutin [177]). $\widetilde{\deg}_\varepsilon(\mathsf{Col}_n^k) = \Omega(\sqrt[3]{(1/2-\varepsilon)\cdot n/k})$ *for any* $0 < \varepsilon < 1/2$ *and* $k|n$.

See also [80] for a direct constructive proof (using Theorem 8.2.4) for the above theorem in the case that $k = 2$.

We will also utilize the Permutation Testing Problem, or PTP for short. This problem, which is closely related to the Collision problem, was defined in [8], which also (implicitly) proved a bound on its one-sided approximate degree.

Definition 8.2.15 (PTP). Given a function $f : [n] \to [n]$ (represented as a string in $\{0,1\}^{n \log n}$),

1. $\mathsf{PTP}_n(f) = 1$ if f is a permutation.
2. $\mathsf{PTP}_n(f) = 0$ if $f(i)$ differs from every permutation on at least $n/8$ values of i.
3. $\mathsf{PTP}_n(f)$ is undefined otherwise.

Theorem 8.2.16 (Implicit in [8]). *For any* $0 < \varepsilon < 1/6$,

$$\deg_\varepsilon^+(\overline{\mathsf{PTP}}_n) = \Omega(n^{1/3})$$

The SZK protocol described for $\overline{\mathsf{Col}}$ works unmodified for PTP as well.

8.3 Hardness Amplification For Approximate Degree

In this section we prove a weaker version of Theorem 8.1.4. Specifically, we show that for any function f with high approximate degree, composing f with GapMaj yields a function with high discrepancy, and hence the resulting function is hard for any PP algorithm in the query model. Similarly, we show that if f has high positive one-sided approximate degree, then composing f with GapAND yields a function with discrepancy. For the full version of Theorem 8.1.4 see the full version of our paper [64].

8.3.1 Notation

For a partial function f, an integer n and a real $\varepsilon \in (1/2, 1]$, we denote $\mathsf{GapMaj}_{n,\varepsilon}(f)$ by F for convenience, where n and ε will always be clear in the context. We also use $x = (x_1, x_2, \ldots, x_n)$ to denote an input to F, where x_i represents the input to the ith copy of f.

The following simple lemma establishes some basic properties of dual witnesses exhibiting the fact that $\widetilde{\deg}_\varepsilon(f) > d$ or $\deg_\varepsilon^+(f) > d$.

Lemma 8.3.1. *Let $f\colon D \to \{0,1\}$ with $D \subseteq \{0,1\}^M$ be a partial function, ε be a real in $[0,1/2)$, and d be an integer such that $\widetilde{\deg}_\varepsilon(f) > d$.*

Let $\mu\colon \{0,1\}^M \to \mathbb{R}$ be a dual witness to the fact $\widetilde{\deg}_\varepsilon(f) > d$ as per Theorem 8.2.4. If f satisfies the stronger condition that $\deg^+_\varepsilon(f) > d$, let μ to be a dual witness to the fact that $\deg^+_\varepsilon(f) > d$ as per Theorem 8.2.6.

We further define $\mu_+(x) := \max\{0, \mu(x)\}$ and $\mu_-(x) := -\min\{0, \mu(x)\}$ to be two non-negative real functions on $\{0,1\}^M$, and μ^i_- and μ^i_+ be the restrictions of μ_- and μ_+ on $f^{-1}(i)$ respectively for $i \in \{0,1\}$. Then the following holds:

- *μ_+ and μ_- have disjoint supports.* (8.2)
- *$\langle \mu_+, p\rangle = \langle \mu_-, p\rangle$ for any polynomial p of degree $\leq d$, so $\|\mu_+\|_1 = \|\mu_-\|_1 = \frac{1}{2}$.* (8.3)
- *$\|\mu^1_+\|_1 > \varepsilon$ and $\|\mu^0_-\|_1 > \varepsilon$. If $\deg^+_\varepsilon(f) > d$, then $\|\mu^1_+\|_1 = 1/2$.* (8.4)

The lemma follows directly from Theorem 8.2.4. We provide a proof below for completeness.

Proof of Lemma 8.3.1. The first two claims follows directly from Theorem 8.2.4 and the definitions of μ_+ and μ_-.

For the third claim, by Theorem 8.2.4, we have

$$\sum_{x\in D} f(x)\cdot\mu(x) - \sum_{x\notin D} |\mu(x)| > \varepsilon.$$

$$\text{Hence, } \|\mu^1_+\|_1 - \|\mu^1_-\|_1 - \sum_{x\notin D} |\mu_-(x)| > \varepsilon.$$

$$\text{This implies that } \|\mu^1_+\|_1 - \|\mu^1_-\|_1 - (0.5 - \|\mu^1_-\|_1 - \|\mu^0_-\|_1) > \varepsilon.$$

$$\text{Hence, } \|\mu^1_+\|_1 - (0.5 - \|\mu^0_-\|_1) > \varepsilon.$$

Therefore, $\|\mu^1_+\|_1 > \varepsilon$, and $(0.5 - \|\mu^0_-\|_1) < 0.5 - \varepsilon$, which means $\|\mu^0_-\|_1 > \varepsilon$.

Finally, the last claim follows directly from Theorem 8.2.6. □

8.3.2 A PP Lower Bound

Here we establish a simpler hardness amplification theorem for $\mathsf{PP}^{\mathsf{dt}}$.

Theorem 8.3.2. *Let $f : D \to \{0,1\}$ with $D \subseteq \{0,1\}^M$ be a partial function, n, d be two positive integers, and $1/2 < \varepsilon < 1$ and $0 < \varepsilon_2 < 1/2$ be two constants such that $2\varepsilon_2 > \varepsilon$. Suppose $\widetilde{\deg}_{\varepsilon_2}(f) > d$. Then*

$$\mathsf{PP}^{\mathsf{dt}}(\mathsf{GapMaj}_{n,\varepsilon}(f)) > \Omega\left\{\min\left(d, (2\varepsilon_2 - \varepsilon)^2 \cdot n\right)\right\}.$$

Proof. For $i \in \{0,1\}$ let $\mu_+, \mu_-, \mu^i_+, \mu^i_-$ be functions whose existence is guaranteed by Lemma 8.3.1, combined with the assumption that $\widetilde{\deg}_{\varepsilon_2}(f) > d$.

In light of Lemma 8.2.8, it suffices to show that $\widetilde{\deg}_{1/2-2^{-T}}(\mathsf{GapMaj}_{n,\varepsilon}(f)) > T$, for $T = \Omega\left\{\min\left(d, (2\varepsilon_2 - \varepsilon)^2 \cdot n\right)\right\}$. We prove this by constructing a dual witness to this fact, as per Theorem 8.2.4.

We first define the following two non-negative functions on $\{0,1\}^{n \cdot M}$:

$$\psi^+(x) := \prod_{i=1}^{n} \mu_+(x_i) \quad \text{and} \quad \psi^-(x) := \prod_{i=1}^{n} \mu_-(x_i).$$

Our dual witness ψ is simply their linear combination:

$$\psi := 2^{n-1} \cdot (\psi^+ - \psi^-).$$

We remark that ψ is precisely the function denoted by ψ_{BT} alluded to in Section 8.1.4. Now we verify that ψ is the dual witness we want.

Proving the ψ has unit ℓ_1-norm. Since μ_+ and μ_- have disjoint supports by Condition (8.2) of Lemma 8.3.1, so does ψ^+ and ψ^-. Therefore $\|\psi\|_1 = 2^{n-1} \cdot (2^{-n} + 2^{-n}) = 1$.

Proving the ψ has pure high degree d. Let $p\colon \{0,1\}^{n \cdot M} \to \mathbb{R}$ be any monomial of degree at most d, and let $p_i\colon \{0,1\}^M \to \mathbb{R}$ be such that $p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_i(x_i)$. Then it holds that

$$\langle \psi^+, p \rangle = \prod_{i=1}^{n} \langle \mu_+, p_i \rangle = \prod_{i=1}^{n} \langle \mu_-, p_i \rangle = \langle \psi^-, p \rangle,$$

where the second equality holds by Condition (8.3) of Lemma 8.3.1.

As a polynomial is a sum of monomials, by linearity, it follows that $\langle \psi, p \rangle = \langle \psi^+, p \rangle - \langle \psi^-, p \rangle = 0$ for any polynomial p with degree at most d.

Proving that ψ has high correlation with F. Define $\mathcal{D}_0 := 2 \cdot \mu_-$ and $\mathcal{D}_1 := 2 \cdot \mu_+$. Note μ_+ and μ_- are non-negative functions with norm $1/2$, so $\mathcal{D}_0$ and $\mathcal{D}_1$ can be thought as distributions on $\{0,1\}^M$. We further define distributions $\mathcal{U}_i$ on $\{0,1\}^{n \cdot M}$ for $i \in \{0,1\}$ as $\mathcal{U}_i := \mathcal{D}_i^{\otimes n}$. Observe that $\mathcal{U}_0 = 2^n \cdot \psi^-$ and $\mathcal{U}_1 = 2^n \cdot \psi^+$ as functions.

Then by Condition (8.4) of Lemma 8.3.1, we have $\Pr_{x \sim \mathcal{D}_1}[f(x) = 1] = 2 \cdot \|\mu_+^1\|_1 > 2\varepsilon_2 > \varepsilon$, and $\Pr_{x \sim \mathcal{D}_0}[f(x) = 0] = 2 \cdot \|\mu_-^0\|_1 > 2\varepsilon_2 > \varepsilon$.

Let D_F denote the domain of F. By the definition of $F = \mathsf{GapMaj}_{n,\varepsilon}(f)$ and a simple Chernoff bound, we have

$$2^n \cdot \sum_{x \in D_F} \psi^+(x) \cdot F(x) = \Pr_{x \sim \mathcal{U}_1}[F(x) = 1] \geq 1 - 2^{-c_1 \Delta^2 \cdot n}, \tag{8.5}$$

where c_1 is a universal constant and $\Delta := 2\varepsilon_2 - \varepsilon$. For brevity, let k denote $c_1 \Delta^2 \cdot n$.

Since $2^n \cdot \|\psi^+\|_1 = 1$, inequality (8.5) further implies that

$$2^n \cdot \sum_{x \notin D_F} \psi^+(x) \leq 2^{-k}.$$

Similarly, we have

$$\Pr_{x \sim \mathcal{U}_0}[F(x) = 0] \geq 1 - 2^{-k},$$

which implies that

$$2^n \cdot \sum_{x \notin D_F} \psi^-(x) \leq 2^{-k}.$$

Putting everything together, we can calculate the correlation between F and ψ as follows:

$$\begin{aligned}
&\sum_{x \in D_F} F(x)\psi(x) - \sum_{x \notin D_F} |\psi(x)| \\
\geq &2^{n-1} \cdot \sum_{x \in D_F} \psi^+(x)F(x) - 2^{n-1} \cdot \left(\sum_{x \notin D_F} \psi^-(x) + \sum_{x \notin D_F} \psi^+(x) \right) \\
\geq &1/2 - 2^{-k-1} - 2^{-k} \\
> &1/2 - 2^{-k+1}.
\end{aligned}$$

Setting $T = \min(d, k-1)$, then we can see that ψ is a dual witness for $\widetilde{\deg}_{1-2^{-T}}(\mathsf{GapMaj}_{n,\varepsilon}(f)) > T$. Clearly $T = \Omega\left\{\min\left(d, (2\varepsilon_2 - \varepsilon)^2 \cdot n\right)\right\}$, which completes the proof. □

8.4 $\mathsf{NISZK}^{\mathcal{O}} \not\subset \mathsf{PP}^{\mathcal{O}}$

In this section we construct an oracle $\mathcal{O}$ such that $\mathsf{NISZK}^{\mathcal{O}} \not\subset \mathsf{PP}^{\mathcal{O}}$. We will use the function $\mathsf{GCol}_n := \mathsf{GapMaj}_{n^{1/4}, 1-\frac{1}{3\log n}}(\mathsf{Col}^{3\log n}_{n^{3/4}})$ to attain the desired oracle separation.

We first show that its complement $\overline{\mathsf{GCol}_n}$ is easy for NISZK by providing a reduction from it to the statistical distance from uniform (SDU) problem. SDU is complete for NISZK and so has an NISZK protocol [127]. We first introduce the problem SDU.

Definition 8.4.1 (Statistical Distance from Uniform (SDU) [127])**.** The promise problem Statistical Distance from Uniform, denoted $\mathsf{SDU} = (\mathsf{SDU}_{\mathsf{YES}}, \mathsf{SDU}_{\mathsf{NO}})$, consisted of

$$\begin{aligned}
\mathsf{SDU}_{\mathsf{YES}} &= \{X : \|X - U\| < 1/n\} \\
\mathsf{SDU}_{\mathsf{NO}} &= \{X : \|X - U\| > 1 - 1/n\}
\end{aligned}$$

where X is a distribution encoded as a circuit outputting n bits, and U is the uniform distribution on n bits, and $\|X - U\|$ denotes the statistical distance between X and U.

Theorem 8.4.2. *There is a* polylog(n)*-time* NISZK *protocol for* $\overline{\mathsf{GCol}_n}$.

Proof. For simplicity, we assume n is a power of 2. We prove this theorem by showing a reduction from $\overline{\mathsf{GCol}_n}$ to an instance of SDU with distributions on $\log n$ bits.

Now, let $m = n^{1/4}$, $k = n^{3/4}$ and $x = (f_1, f_2, \ldots, f_m)$ be an input to GCol_n, where each f_i is interpreted as a function from $[k] \to [k]$. We construct the distribution $\mathcal{D}(x)$ as follows: to generate a sample from $\mathcal{D}(x)$, we pick a pair $(i, j) \in [m] \times [k]$ at uniformly random, and output the sample $(i, f_i(j))$. Clearly $\mathcal{D}(x)$ is polylog(n)-time preparable.

Now we show this is a valid reduction. Let $\mathcal{U}$ be the uniform distribution on $[m] \times [k]$ and $\mathcal{U}_k$ be the uniform distribution on $[k]$. For a function $f : [k] \to [k]$, let $\mathcal{D}_f$ be the distribution obtained by outputting $f(i)$ for an index $i \sim \mathcal{U}_k$. Then we can see $\mathcal{D}(x) = \frac{1}{m}\sum_{i=1}^{m} \{i\} \times \mathcal{D}_{f_i}$.

When $\overline{\mathsf{GCol}_n}(x) = 1$, we have

$$\|\mathcal{D}(x) - \mathcal{U}\| = \frac{1}{m}\sum_{i=1}^{m} \|\mathcal{U}_k - \mathcal{D}_{f_i}\| \leq \frac{1}{3\log n} < \frac{1}{\log n}.$$

Here, the first inequality holds because at least a $1 - \frac{1}{3\log n}$ fraction of f_i's are permutations, which implies that $\|\mathcal{U}_k - \mathcal{D}_{f_i}\| = 0$.

When $\overline{\mathsf{GCol}_n}(x) = 0$, we have

$$\|\mathcal{D}(x) - \mathcal{U}\| = \frac{1}{m}\sum_{i=1}^{m} \|\mathcal{U}_k - \mathcal{D}_{f_i}\| \geq \left(1 - \frac{1}{3\log n}\right) \cdot \left(1 - \frac{1}{3\log n}\right) > 1 - \frac{1}{\log n}.$$

Here, the first inequality holds because at least a $1 - \frac{1}{3\log n}$ fraction of f_i's are $3\log n$-to-1, which implies that $\|\mathcal{U}_k - \mathcal{D}_{f_i}\| = 1 - \frac{1}{3\log n}$.

Putting everything together, we have shown $\mathcal{D}(x)$ that is a valid reduction to SDU. This completes the proof.

□

Then by a straightforward application of Theorem 8.3.2, we can show GCol_n is hard for any PP algorithm.

Theorem 8.4.3. $\mathsf{PP}^{\mathsf{dt}}(\mathsf{GCol}_n) = \Omega(n^{1/4}/\log n)$.

Proof. Observe that $\widetilde{\deg}_{1/2-\frac{1}{50\log n}}(\mathsf{Col}^{3\log n}_{n^{3/4}}) = \Omega(n^{1/4}/\log^{2/3} n)$ by Theorem 8.2.14. Applying Theorem 8.3.2 with $a = \dfrac{1 - 2\cdot\frac{1}{50\log n}}{2\cdot\frac{1}{50\log n}} = 25\log n - 1$ (recall $a = \dfrac{2\varepsilon_2}{1-2\varepsilon_2}$ in Theorem 8.3.2), we have that

$$\begin{aligned}
&\mathsf{PP}^{\mathsf{dt}}\left(\mathsf{GapMaj}_{n^{1/4},1-\frac{1}{3\log n}}(\mathsf{Col}^{3\log n}_{n^{3/4}})\right) \\
\geq &\deg_{\pm}\left(\mathsf{GapMaj}_{n^{1/4},1-\frac{1}{3\log n}}(\mathsf{Col}^{3\log n}_{n^{3/4}})\right) && \text{(by Lemma 8.2.9)} \\
\geq &\min\left\{\left(1 - \left(1 + \frac{10}{a}\right)\cdot\left(1 - \frac{1}{3\log n}\right)\right)\cdot n^{1/4} - 4, \widetilde{\deg}_{1/2-\frac{1}{50\log n}}(\mathsf{Col}^{3\log n}_{n^{3/4}})\right\} \\
\geq &\Omega(n^{1/4}/\log n).
\end{aligned}$$

□

Now Theorem 8.1.1 from the introduction follows from standard diagonalization methods and the observation that PP is closed under complement.

8.5 Limitations on Perfect Zero Knowledge Proofs (Proof of Theorem 8.1.2)

In this section, we study the limitations of perfect zero knowledge in the presence of oracles.

Definition 8.5.1 (Honest Verifier Perfect Zero Knowledge). A promise problem $L = (L_Y, L_N)$ is in HVPZK if there exist a tuple of Turing machines (P, V, S), where the *verifier* V and *simulator* S run in probabilistic polynomial time, satisfying the following:

- (P, V) is an interactive proof for L with negligible completeness and soundness errors.
- $S(x)$ is allowed to fail (by outputting $\perp$), but with probability at most $1/2$.
- For any $x \in L$, let $\hat{S}(x)$ denote the distribution of $S(x)$ conditioned on it not failing. Then,

$$\|\hat{S}(x) - (P, V)(x)\| = 0$$

The class PZK is defined similarly but, as in the case of SZK, with the additional stipulation that for any verifier that deviates from the protocol, there is a simulator that simulates the prover's interaction with that verifier. It is easy to see that PZK $\subseteq$ HVPZK in the presence of any oracle. (The definitions of these classes in the presence of an oracle are the same, except that P, V, and S all have access to the oracle.)

Note that the probability of failure of the simulator can be made negligible (in $|x|$) by repeating it a polynomial number of times and taking its output to be that from the first time that it succeeds. We use this implicitly in the rest of our development. The variant where the simulator is not allowed to fail is called Super-Perfect Zero Knowledge by Goldreich and Teichner [129]. There (and also elsewhere), this definition is considered to be "oversimplified" as such proof systems are not known for problems outside BPP. However, in the setting of honest verifiers with small but non-zero completeness error, the class thus defined turns out to be equal to HVPZK [129]. While sometimes these classes are defined with the requirement of perfect completeness in the zero knowledge proofs, note that defining them as above only makes our results stronger – requiring perfect completeness can only make HVPZK smaller, and our oracle separation between PZK and coPZK continues to hold when both these classes are defined with perfect completeness.

8.5.1 A Preliminary Lemma

We will need the following lemma in the proof of the theorems that follow.

Lemma 8.5.2. *There is an oracle Turing Machine M_2 that is such that when given sample access to two distributions p and q, M_2 uses two samples and,*

$$\Pr[M_2^{p,q} \text{ accepts}] = \frac{1}{2} + \frac{\|p - q\|_2^2}{8}$$

Proof. $M_2^{p,q}$ behaves as follows:

1. With probability $\frac{1}{4}$, sample y_1, y_2 from p.
   - If $y_1 = y_2$, accept with probability 1.
   - Else, accept with probability $\frac{1}{2}$.
2. With probability $\frac{1}{4}$, do the same with samples from q.

3. With probability $\frac{1}{2}$, sample y_1 from p and y_2 from q.

   - If $y_1 = y_2$, reject with probability 1.
   - Else, accept with probability $\frac{1}{2}$.

$$\begin{aligned}\Pr[M_2^{p,q} \text{ accepts}] &= \frac{1}{4}\left[(1 - \|p\|_2^2)\frac{1}{2} + \|p\|_2^2\right] + \frac{1}{4}\left[(1 - \|q\|_2^2)\frac{1}{2} + \|q\|_2^2\right] \\ &\quad + \frac{1}{2}\left[(1 - \langle p, q\rangle)\frac{1}{2} + \langle p, q\rangle \cdot 0\right] \\ &= \frac{1}{2} + \frac{\|p - q\|_2^2}{8}\end{aligned}$$

□

8.5.2 Showing HVPZK $\subseteq$ PP Relative to Any Oracle

The first step in our proof of Theorem 8.1.2 is to show that HVPZK is contained in PP in a relativizing manner.

Theorem 8.5.3. HVPZK $\subseteq$ PP. *Further, this is true in the presence of any oracle.*

Proof. Let L be a language with an HVPZK proof system (P, V, S). We will show how to decide membership in L in PP. Fix any input length n, and let the number of messages in the proof system for any input of this length be m, and the length of each message be ℓ (these are without loss of generality). Also suppose that the first message is always sent by the verifier. Let the number of random bits used by V on an input of length n be v, and the number of random bits used by S be s.

For any $x \in \{0,1\}^n$, we write the output of the simulator S on input x using randomness r as $S(x;r) = (R_V(x;r), T_1(x;r), \ldots, T_m(x;r))$, where R_V is the simulated randomness of the verifier, and T_i is the simulated ith message in the protocol. Let S_i denote S truncated at T_i. Denote by V_S the verifier simulated by S, and by P_S the simulated prover, both conditioned on the simulator not failing.

Claim 15. *An input x is in L if and only if the following three conditions are satisfied:*

1. *V_S on input x behaves like the actual verifier V. This involves the following:*
   - *$R_V(x)$, conditioned on not being $\perp$, is distributed uniformly over $\{0,1\}^v$.*
   - *For any non-failing transcript $(r_V, t_1, \ldots, t_m)$ output by $S(x)$, the verifier's responses in $(t_1, \ldots, t_m)$ are consistent with what V would have sent when using r_V as randomness.*
2. *P_S on input x is a valid prover.*
   - *This means that the distribution of the prover's simulated messages $(T_{2i}(x))$ should depend only on the messages in the transcript so far $(T_1(x), \ldots, T_{2i-1}(x))$, and should be independent of the verifier's simulated randomness $(R_V(x))$.*
3. *$S(x)$ is an accepting transcript with probability at least $3/4$.*

For any $x \in L$, the transcript of the actual protocol satisfies the above properties, and so does the simulation, since it is perfect conditioned on not failing.

The other direction follows on noting that if all three conditions are satisfied for some x, then P_S is a prover strategy that convinces the actual verifier V that $x \in L$. By the soundness of the (P, V) proof system, this can only happen if x is indeed in L.

So to decide the membership of x in L in PP, it is sufficient to be able to decide each of the above three properties of $S(x)$ in PP (since PP is closed under conjunction [54]). Of these, property (3) is easily seen to be decidable in BPP, and hence in PP.

Lemma 8.5.2 says, in particular, that testing whether two polynomial-time-sampleable distributions p and q are identical can be done in PP. Let $U_S(x)$ be the distribution sampled by first running $S(x)$, outputting $\perp$ if it fails, and a uniform sample from $\{0,1\}^v$ if it doesn't. The first check on V_S is the same as checking whether $R_V(x)$ is identical to $U_S(x)$. The other check required on V_S is a coNP statement, and hence can be done in PP.

Let M_2 be the TM from Lemma 8.5.2. To check that P_S is a valid prover, consider the TM – call it M_P – that works as follows on input x.

1. Select $i \in \left[-1, \frac{m}{2}\right]$ at random.
2. If $i = 0$, run M_2 on the distributions $R_V(x)$ and $U_S(x)$.
3. If $i = -1$, check the consistency of transcripts produced by $S(x)$ with the simulated randomness.
   - This is done by selecting $r_S \in \{0,1\}^s$, and running $S(x; r_S)$ to get $(r_V, t_1, \ldots, t_m)$.
   - If this transcript is failing or consistent, accept with probability $1/2$, else with probability 1.
4. Else, select at random $t_1, \ldots, t_{2i} \in \{0,1\}^\ell$, $r_V^1, r_V^2 \in \{0,1\}^v$, and $r_S^1, r_S^2 \in \{0,1\}^s$.
5. If $S(x; r_S^1)$ does not have $(r_V^1, t_1, \ldots, t_{2i-1})$ as a prefix or $S(x; r_S^2)$ does not have $(r_V^2, t_1, \ldots, t_{2i-1})$ as a prefix, accept with probability $1/2$.
6. Let p be the distribution over $\{0,1\}$ such that $p(1) = \Pr[S_{2i}(x) = (r_V^1, t_1, \ldots, t_{2i})]$, and q be the same but with r_V^2 instead of r_V^1.
7. Run M_2 on the distributions p and q.

Claim 16. *$M_P(x)$ accepts with probability at most $\frac{1}{2}$ if and only if V_S is a valid verifier and P_S is a valid prover on input x.*

Suppose V_S is a valid verifier and P_S is a valid prover on input x. If M_P selects $i = 0$ or $i = -1$, then it accepts with probability $\frac{1}{2}$ because V_S is a valid verifier.

If $i \notin \{-1, 0\}$, and M_P picks $r_V^1, r_V^2, t_1, \ldots, t_{2i}$. If this fails the check in step 5, then M_P again accepts with probability $1/2$. If this does not happen and $r_V^1, r_V^2, t_1, \ldots, t_{2i-1}$ are in the support of $S_{2i-1}(x)$,

$$\begin{aligned}
&\Pr[S_{2i}(x) = (r_V^1, t_1, \ldots, t_{2i})] \\
&\quad = \Pr[S_{2i-1}(x) = (r_V^1, t_1, \ldots, t_{2i-1})] \Pr[T_{2i}(x) = t_{2i} \mid S_{2i-1}(x) = (r_V^1, t_1, \ldots, t_{2i-1})] \\
&\quad = \Pr[S_{2i-1}(x) = (r_V^1, t_1, \ldots, t_{2i-1})] \Pr[T_{2i}(x) = t_{2i} \mid S_{2i-1}(x) = (r_V^2, t_1, \ldots, t_{2i-1})]
\end{aligned}$$

where the second equality is because P_S is a valid prover, so its responses do not depend on the simulated randomness of the verifier. We can write the first term in the product above as:

$$\begin{aligned}
&\Pr[S_{2i-1}(x) = (r_V^1, t_1, \ldots, t_{2i-1})] \\
&\quad = \Pr[S_{2i-2}(x) = (r_V^1, t_1, \ldots, t_{2i-2})] \Pr[T_{2i-1}(x) = t_{2i-1} \mid S_{2i-2}(x) = (r_V^1, t_1, \ldots, t_{2i-2})] \\
&\quad = \Pr[S_{2i-2}(x) = (r_V^1, t_1, \ldots, t_{2i-2})]
\end{aligned}$$

where the second equality is because V_S is a valid verifier and is deterministic once R_V is fixed, and step 5 was there precisely to check that this probability is non-zero.

Now starting from the fact that $\Pr[S_0(x) = r_V^1] = \Pr[S_0(x) = r_V^2]$, and using the above relationships, we can inductively prove that $\Pr[S_{2i}(x) = (r_V^1, t_1, \ldots, t_{2i})] = \Pr[S_{2i}(x) = (r_V^2, t_1, \ldots, t_{2i})]$. This implies that the call to M_2 in step 7 of M_P accepts with probability $1/2$, as the distributions p and q there are identical. So in all cases, M_P accepts with probability $1/2$.

To prove the converse, we start by noting that each branch of M_P always accepts with probability $1/2$ or more. So even if one of the branches accepts with probability strictly more than $1/2$, the acceptance probability of M_P as a whole will be strictly more than $1/2$.

Now suppose V_S is not a valid verifier. Then M_P would accept with probability strictly more than $1/2$ because either $i = 0$ or $i = -1$ would accept with probability more than $1/2$.

The remaining case is where V_S is a valid verifier but P_S is not a valid prover. This means that at some point the distribution of P_S's responses depended on the simulated verifier's randomness. Specifically, there must exist an $i \in [m/2]$ and $r_V^1, r_V^2, t_1, \ldots, t_{2i}$ such that $(\{r_V^1, r_V^2\}, t_1, \ldots, t_{2i-1})$ are in the support of $S_{2i-1}(x)$ and:

$$\begin{aligned}
&\Pr[T_{2i}(x) = t_{2i} \mid S_{2i-1}(x) = (r_V^1, t_1, \ldots, t_{2i-1})] \\
&\neq \Pr[T_{2i}(x) = t_{2i} \mid S_{2i-1}(x) = (r_V^2, t_1, \ldots, t_{2i-1})]
\end{aligned}$$

For this r_V^1 and r_V^2, let i_0 be the least i such that there exist $t_1, \ldots, t_{2i_0}$ where such an inequality holds. i_0 being the smallest such i implies, by the same induction arguments above and the validity of V_S as a verifier, that:

$$\Pr[S_{2i_0-1}(x) = (r_V^1, t_1, \ldots, t_{2i_0-1})] = \Pr[S_{2i_0-1}(x) = (r_V^2, t_1, \ldots, t_{2i_0-1})]$$

Putting the above two relations together, we get:

$$\Pr[S_{2i_0}(x) = (r_V^1, t_1, \ldots, t_{2i_0})] \neq \Pr[S_{2i_0}(x) = (r_V^2, t_1, \ldots, t_{2i_0})]$$

So when M_P chooses $i = i_0$ and these values of r_V^1, r_V^2 and $t_1, \ldots, t_{2i_0}$, it will accept with probability strictly greater than $1/2$, and so it will do so overall as well. This proves Claim 16.

Due to the fact that PP is closed under complement and Claim 16, we have now established that the conditions in Claim 15 can be checked in PP. And so by Claim 15, L can be decided in PP. It is also easy to see that this proof still works relative to any oracle, as it only makes black-box use of S. □

The following theorem follows immediately from Corollary 8.1.1 and Lemma 8.5.3.

Theorem 8.5.4. *There is an oracle $\mathcal{O}$ such that* $\mathsf{NISZK}^{\mathcal{O}} \not\subseteq \mathsf{HVPZK}^{\mathcal{O}}$. *Consequently,* $\mathsf{SZK}^{\mathcal{O}} \not\subseteq \mathsf{PZK}^{\mathcal{O}}$ *and* $\mathsf{NISZK}^{\mathcal{O}} \not\subseteq \mathsf{NIPZK}^{\mathcal{O}}$.

8.5.3 A Relativized Separation of PZK and coPZK

Theorem 8.5.5. *There is an oracle $\mathcal{O}$ such that* $\mathsf{PZK}^{\mathcal{O}} \neq \mathsf{coPZK}^{\mathcal{O}}$.

Proof. In order to prove Theorem 8.5.5, we first show that HVPZK is closed under "composition" with GapAND.

Lemma 8.5.6. *Let $f : D \to \{0,1\}$ with $D \in \{0,1\}^M$ be a partial function and n be a positive integer, $1/2 < \varepsilon < 1$ be a constant. If f has a* polylog(M)*-time* HVPZK *protocol, then* $\mathsf{GapAND}_{n,\varepsilon}(f)$ *has a* polylog(nM)*-time* HVPZK *protocol.*

Proof. For convenience, denote $\mathsf{GapAND}_{n,\varepsilon}(f)$ by g. Given an HVPZK protocol (P, V, S) for f, we will construct an HVPZK protocol (P', V', S') for g. Given an input $x = (x_1, \ldots, x_n)$ for g, V' selects, say, $\log^2(n)$ values of $i \in [n]$, and P' and V' run the interactive protocol (P, V) on each of the corresponding x_i's independently. V' accepts if and only if (P, V) accepts on all these x_i's. Completeness and soundness follows easily from standard arguments and the definition of g.

On a similar input, the simulator S' simply selects the same number of i's, and runs S on the corresponding x_i's. Since in a YES instance all of the x_i's are such that $f(x_i) = 1$, S simulates the transcripts of (P, V) on all of these exactly, except with a negligible probability when it fails on one or more of these. Hence S' simulates (P', V') exactly as well, again failing only with a negligible probability. □

We will need the following implication of the constructions in [98].

Lemma 8.5.7 (Implied by [98])**.** *Any partial Boolean function that has a* polylog(n)*-time* NIPZK *protocol also has a* polylog(n)*-time* PZK *protocol*[9].

We will use the function PTP_n (cf. Definition 8.2.15) to establish our separation. The following are immediate consequences of Theorems 8.2.16 and 8.1.4 and Lemma 8.2.8.

Corollary 8.5.8. $\mathsf{PP}^{\mathsf{dt}}(\mathsf{GapAND}_{n,7/8}(\overline{\mathsf{PTP}}_n)) = \Omega(n^{1/3})$

Lemma 8.5.9. PTP_n *has a* polylog(n)*-time* PZK *protocol.*

Proof. We will show this by presenting a polylog(n)-time NIPZK protocol for PTP_n and invoking Lemma 8.5.7. The protocol is very similar to the one described in Section 8.2.5. Given a function $f : [n] \to [n]$ as input, an $r \in [n]$ is chosen at random using the common random string. P is then supposed to send an x to V such that $f(x) = r$. V accepts if this is true. Completeness, soundness and perfect zero-knowledge are all easily argued using the definition of PTP_n. □

Now we have everything we need to prove Theorem 8.5.5. Suppose $\mathsf{PZK}^{\mathcal{O}} = \mathsf{coPZK}^{\mathcal{O}}$ with respect to all oracles $\mathcal{O}$. This implies that any language that is in polylog(n)-time PZK is also in polylog(n)-time coPZK, and vice versa – if this were not true for some language, then we would be able to use that language to construct an oracle that separates the two classes by diagonalization. In particular, this hypothesis and Lemma 8.5.9 imply that $\overline{\mathsf{PTP}}_n$ has a polylog(n)-time PZK (and hence HVPZK) protocol. Then, by Lemma 8.5.6, $\mathsf{GapAND}_{n,7/8}(\overline{\mathsf{PTP}}_n)$ has a polylog(n)-time HVPZK protocol.

[9] This is a nontrivial fact as NIPZK is defined such that both parties have access to a shared random string, while PZK is not. So NIPZK is trivially in HVPZK (the honest verifier can generate a random string and send it to the prover), but is not obviously in PZK.

This fact, along with the lower bound in Corollary 8.5.8, can be used to construct an oracle separating HVPZK from PP by standard diagonalization. But by Lemma 8.5.3, such an oracle cannot exist. So there has to be some oracle separating PZK and coPZK, completing the proof of Theorem 8.5.5. □

An argument identical to the proof of Theorem 8.5.5 (without the need to invoke Lemma 8.5.7) shows that the same oracle separates NIPZK and coNIPZK, as well as HVPZK and coHVPZK.

Theorem 8.5.10. *The oracle $\mathcal{O}$ witnessing Theorem 8.5.5 also satisfies* $\mathsf{NIPZK}^{\mathcal{O}} \neq \mathsf{coNIPZK}^{\mathcal{O}}$, *as well as* $\mathsf{HVPZK}^{\mathcal{O}} \neq \mathsf{coHVPZK}^{\mathcal{O}}$.

Combining Theorems 8.5.3, 8.5.4, 8.5.5, and 8.5.10 yields Theorem 8.1.2 from Section 8.1.2.

8.6 Consequences for Polarization

8.6.1 Introduction to Polarization and Summary of Our Results

A polarization algorithm is an algorithm that is given black-box sampling access to two distributions, and outputs two new distributions that are either extremely close in total variation distance (if they were initially somewhat close) or extremely far in total variation distance (if they were originally somewhat far). In this section we describe how our oracle separation between SZK and PP implies lower bounds on polarization algorithms. In particular we show black-box polarization algorithms are limited in how close they can push the statistical difference to 0 or 1 relative to the number of bits in the output distribution.

The concept of polarization first arose in work of Sahai and Vadhan [211]. In their work, Sahai and Vadhan showed that the statistical difference problem is complete for the class SZK. The statistical distance problem is formulated as follows: Let $P_b(x)$ be poly-sized classical circuits. Let D_b be the distribution on $\{0,1\}^n$ induced by inputting a uniformly random input x to $P_b(x)$. The statistical difference problem is, given circuits P_0 and P_1, determine if either $||D_0 - D_1|| \leq 1/3$ or if $||D_0 - D_1|| \geq 2/3$, promised one is the case. Here $||D_0 - D_1||$ indicates the total variation distance between the distributions D_0 and D_1.

In their paper, Sahai and Vadhan also showed a remarkable property of the statistical difference problem – namely that the constants $1/3$ and $2/3$ in the Statistical Difference problem can be amplified to be exponentially close to 0 and 1 [211]. This property is not immediately obvious, because it cannot be obtained by simply repeatedly sampling from D_0 and D_1. Nevertheless, they showed the following: given black-box distributions D_0 and D_1, and a number k expressed in unary, then in polynomial time one can sample from distributions D_0' and D_1' (using polynomially many samples from D_0 and D_1) such that, if $||D_0 - D_1|| \leq 1/3$, then $||D_0' - D_1'|| \leq \varepsilon$ and if $||D_0 - D_1|| \geq 2/3$, then $||D_0' - D_1'|| \geq 1 - \varepsilon$, where $\varepsilon = 2^{-k}$. Hence without loss of generality, one can assume that the distributions in the statistical different problem are exponentially close to 0 or 1; their transformation "polarizes" the distributions to be either very close or very far from one another. This is known as the Polarization Lemma, and is a key part of the proof that Statistical Difference is SZK-complete[10].

[10] In statistical zero-knowledge proof systems, the verifier must be able to simulate the honest prover to negligibly small (1/superpoly) total variation distance. The ability to polarize distributions allows the statistical difference problem to have this property.

Given this fundamental result, it is natural to ask whether or not one can improve the parameters of the Polarization Lemma. For instance, Sahai and Vadhan noted in their paper that their algorithm could only polarize distributions under the promise $||D_0 - D_1|| > \alpha$ or $||D_0 - D_1|| < \beta$ in the case that $\alpha^2 > \beta$. So their algorithm can polarize $\alpha = 2/3$ and $\beta = 1/3$, but not $\alpha = 5/9$ and $\beta = 4/9$. A natural question is whether or not this limitation could be removed. Holenstein and Renner answered this question in the negative for certain types of black-box polarization [154]. In particular, they showed that any form of black-box polarization which works by drawing strings $b, c \in \{0,1\}^\ell$, and then sets $D'_0 = D_{b_1} \otimes ... \otimes D_{b_\ell}$ and $D'_1 = D_{c_1} \otimes ... \otimes D_{c_\ell}$ cannot polarize in the case where $\alpha^2 < \beta$. As Sahai and Vadhan's polarization algorithm took this form, this was strong evidence that this limitation was fundamental. Note, however, that it remains open to show that polarization cannot occur when $\alpha^2 < \beta$ using *arbitrary* black-box algorithms. For instance, one could feed the random outputs of of D_0 back into the circuit for D_1 in order to help polarize the distributions. While it is not clear how these sorts of operations could help one polarize, it is difficult to rule out the possibility that such operations might lead to a stronger polarization algorithm.

In this section we consider different parameters of the Polarization Lemma - namely how small can the security parameter ε be relative to the size of the range of the output distributions. For example, if one is given distributions D_0 and D_1 over n-bit strings with total variation distance $> 2/3$ or $< 1/3$, then can one create distributions D'_0 and D'_1 over n'-bit string such that the total variation distance is $\leq \varepsilon$ or $\geq 1 - \varepsilon$ where $\varepsilon = 2^{-n'}$, or $2^{-n'^2}$? At first it might appear the answer to the above question is trivially yes - because one can simply set $k = -n^2$ (or $k = n^c$ for any constant c) and run the Polarization Lemma. However this does not work because the Polarization Lemma increases the size of the domains of the distributions as it polarizes; in other words n' is some polynomial function of n and k. By tweaking the parameters of the Polarization Lemma slightly [211], one can polarize distributions on n bits to distributions on $n' = \text{poly}(n)$ bits which are polarized to roughly $\varepsilon \approx 2^{-\sqrt{n'}}$. However, it seems difficult to do better than $\varepsilon = 2^{-\sqrt{n'}}$ using the proof techniques of Sahai and Vadhan [211]. This is because their proof alternates between two lemmas, one which total variation distance towards 1 in the case the distributions are far apart, and another which pushes the total variation distance towards zero in the case the distributions are close. In order to make the distributions 2^{-k}-close or $1 - 2^{-k}$-far, one must apply both lemmas, each of which increases the number of bits output by the distribution by a factor of k. Hence using Sahai and Vadhan's Lemma with $k = n^c$, the best one can achieve are distributions on $n' = n^{2c+1}$ bits which are either 2^{-n^c}-close or $(1 - 2^{-n^c})$-far. For large constant c this gives $\varepsilon \approx 2^{-\sqrt{n'}}$. It seems difficult to improve their lemma further using the techniques of their paper.

A natural question is therefore: what is the smallest value of ε that one can achieve relative to the size of the output distributions n'? In this section, we show that if ε can be made very small relative to n', then that would place $\mathsf{SZK}^{\mathcal{O}} \subseteq \mathsf{PP}^{\mathcal{O}}$ (and even $\mathsf{SZK}^{\mathcal{O}} \subseteq \mathsf{BPP}_{\mathsf{path}}{}^{\mathcal{O}}$) for all oracles $\mathcal{O}$. Therefore, as a corollary of our main result, ε cannot be made very small by any poly-time black-box polarization algorithm. More specifically, we achieve a lower bound of $\varepsilon > 2^{-n'/2-1}$ for any poly-time polarization algorithm.

More specifically, we prove two theorems showing that a stronger version of polarization places SZK in PP relative to all oracles. Therefore, a stronger polarization algorithm cannot exist as a corollary of Theorem 8.1.1 - which implies Theorem 8.1.3.

Theorem 8.6.1. *Suppose that there is an algorithm running in poly(n) time, which given black box distributions D_0, D_1 on strings of length n which obey either $|D_0 - D_1| < 1/3$*

or $|D_0 - D_1| > 2/3$, *produces two output distributions* D'_0 *and* D'_1 *on strings of length* $n' = poly(n)$ *such that either* $|D'_0 - D'_1| < \varepsilon$ *(in the first case) or* $|D'_0 - D'_1| > 1 - \varepsilon$ *(in the second case) where* $\varepsilon \leq 2^{-n'/2-1}$. *Then* $\mathsf{SZK}^{\mathcal{O}} \subseteq \mathsf{PP}^{\mathcal{O}}$ *for all oracles* $\mathcal{O}$.

Theorem 8.6.2. *Suppose that there is an algorithm running in* $poly(n)$ *time, which given black box distributions* D_0, D_1 *on strings of length* n *which obey either* $|D_0 - D_1| < 1/3$ *or* $|D_0 - D_1| > 2/3$, *produces two output distributions* D'_0 *and* D'_1 *on strings of length* $n' = poly(n)$ *such that either* $|D'_0 - D'_1| < \varepsilon$ *(in the first case) or* $|D'_0 - D'_1| > 1 - \varepsilon$ *(in the second case) where* $\varepsilon \leq 2^{-2n'/3-1}$. *Then* $\mathsf{SZK}^{\mathcal{O}} \subseteq (\mathsf{BPP}_{\mathsf{path}})^{\mathcal{O}}$ *for all oracles* $\mathcal{O}$.

Therefore as a corollary of Theorem 8.1.1, there do not exist poly-time polarization algorithms achieving $\varepsilon = 2^{-n'/2-1}$. In fact one could have achieved such a lower bound even if one had merely given an oracle separation between SZK and $\mathsf{BPP}_{\mathsf{path}}$. It remains open to close the gap between our lower bound of $\varepsilon = 2^{-n'/2-1}$ and the upper bound of $\varepsilon = 2^{-n'^{1/2+\delta}}$ for any $\delta > 0$ given by Sahai and Vadhan [211].

Note that if one wishes to prove our lower bounds for Holenstein-Renner style Polarization only, then there is a more direct proof of this fact using Fourier analysis. We provide this simplified proof at the end of this section to facilitate understanding of why this sort of polarization is impossible. But note that this result is subsumed by Theorems 8.6.1 and 8.6.2.

The proof of Theorem 8.6.1 is relatively straightforward. Suppose one can polarize to $\varepsilon' \ll 2^{-n'/2}$. Then the output distributions now have a promise on the ℓ_2 distance between the output distributions - in particular the ℓ_2 distance between them is more or less than some (exponentially small) threshold. It is easy to decide this problem PP - this is because the ℓ_2 distance square is a degree-two polynomial in the output probabilities. To see this, say you're trying to determine if $S = \sum_{x \in \{0,1\}^n} (D'_0(x) - D'_1(x))^2$ is more or less than some threshold t, consider the following algorithm: pick at random x, pick a random number 1,2,3 or 4. If the number is 1 (respectively 4) sample two samples from D'_0 (respectively D'_1) and accept if they both give output x, otherwise output accept/reject using a 50-50 coin flip. If the number is 2 or 3 sample one sample from D'_0 and D'_1 and reject iff they collide, otherwise output a 50-50 coin flip. The probability this machine accepts is $1/2 + S/2$ - which is more more or less than a known threshold $(1 + t)/2$. Therefore by correcting the bias of the machine with an initial coin flip, this is a PP algorithm to decide the problem. In short, deciding thresholds for the ℓ_2 norm is easy for PP because it is a low-degree polynomial, while deciding thresholds for the ℓ_1 norm is hard for PP because the ℓ_1 norm is not a low degree polynomial.

On the other hand, the proof of Theorem 8.6.2 is involved – it works by examining the algorithms from Chapter 7 showing that certain modified versions of quantum mechanics can be used to solve SZK-hard problems in polynomial time [13]. These algorithms are not based on postselection (otherwise they would place $\mathsf{SZK} \subseteq \mathsf{PostBQP} = \mathsf{PP}$ for all oracles, a contradiction with our main result). However, it turns out that if one has a very strong polarization lemma, then one can turn them into postselected quantum algorithms (and even postselected classical algorithms) for statistical difference. We include this proof below.

8.6.2 Proof of Theorem 8.6.2

Proof of Theorem 8.6.2. To prove the theorem, suppose that the statistical difference problem is SZK-hard for distributions on N bits which are either ε-close or $(1 - \varepsilon)$-far, where

$\varepsilon = o(2^{-2N/3})$. We will give a $\mathsf{BPP}_{\mathsf{path}}$ algorithm to solve this problem, using the characterization that $\mathsf{BPP}_{\mathsf{path}} = \mathsf{postBPP}$. The algorithm is inspired by Aaronson, Bouland, Fitzsimon, and Lee's proof that $\mathsf{SZK} \subseteq \mathsf{naCQP}$ given in [13]. We thank Tomoyuki Morimae and Harumichi Nishimura for helpful discussions on this topic.

The algorithm is as follows: flip three coins b_1, b_2, b_3, and draw independent samples y_1, y_2, y_3 from the distributions $D_{b_1}, D_{b_2}, D_{b_3}$, respectively. Postselect on the condition that $y_1 = y_2 = y_3$. Output that the distributions are far apart if $b_1 = b_2 = b_3$, and otherwise output that the distributions are close.

If $\varepsilon = 0$, then clearly this algorithm is correct. In the case the distributions are far apart, they have disjoint support, which implies the values b_i must be identical, so in this case the algorithm has zero probability of error. In the case the distributions are close, they are identical, so the string $b_1b_2b_3$ is uniformly random after postselection, so the algorithm errs with probability $1/4$. Note that the correctness of this algorithm in the case $\varepsilon = 0$ doesn't tell us anything new in structural complexity, because in the $\varepsilon = 0$ case, the problem is in NP (as a witness to the fact the distributions are identical, simply provide x_0, x_1 such that $P_0(x_0) = P_1(x_1)$), and hence is obviously in $\mathsf{BPP}_{\mathsf{path}}$ and in PP as well.

We now claim that if $\varepsilon = o(2^{-2N/3})$, then this algorithm still works. Note that our choice of ε is asymptotically tight for our algorithm; if $\varepsilon = \Omega(2^{-2N/3})$, then there is a simple counterexample which foils the algorithm [11]. To show that the algorithm works, we'll show two things. First, if the distributions are ε-close for this small ε, then we'll show that as $n \to \infty$, then $\hat{b}$'s value approaches the uniform distribution over all 8 possible output strings. Therefore for sufficiently large n, the algorithm is correct. On the other hand, if the distributions are $1 - \varepsilon$-far, we'll show the algorithm is correct with high probability.

Let's first handle the case in which the distributions are ε-close. Let $\hat{b} \in \{0,1\}^n$ be the random variable corresponding to the output of $b_1b_2b_3$. Let $D_b(y)$ denote the probability that distribution D_b outputs y. Let S be the event that $y_1 = y_2 = y_3$, and let $S(y)$ be the event that $y_1 = y_2 = y_3 = y$. By Bayes' rule, we have that

$$\begin{aligned}
\Pr[\hat{b} = b_1b_2b_3|S] &= \frac{\Pr[S|\hat{b} = b_1b_2b_3]\Pr[\hat{b} = b_1b_2b_3]}{\Pr[S]} \\
&= \frac{\sum_{y\in\{0,1\}^n} \Pr[S(y)|\hat{b} = b_1b_2b_3]\frac{1}{8}}{\sum_{y\in\{0,1\}^n} \Pr[S(y)]} \\
&= \frac{\sum_{y\in\{0,1\}^n} D_1(y)^{w(\hat{b})}D_0(y)^{3-w(\hat{b})}\frac{1}{8}}{\sum_{y\in\{0,1\}^n} \Pr[S(y)]}
\end{aligned}$$

where $w(\hat{b})$ is the Hamming weight of $\hat{b}$.

Hence we have that

$$\frac{\Pr[\hat{b} = b_1b_2b_3|S]}{\Pr[\hat{b'} = b'_1b'_2b'_3|S]} = \frac{\sum_{y\in\{0,1\}^n} D_1(y)^{w(\hat{b})}D_0(y)^{3-w(\hat{b})}}{\sum_{y\in\{0,1\}^n} D_1(y)^{w(\hat{b'})}D_0(y)^{3-w(\hat{b'})}}$$

[11] Let D_0 be a uniform distribution, and let D_1 be the distribution which places an ε amount of weight on a single item x, while the remaining weight is spread uniformly on the remaining elements. These distributions are ε-close in total variation distance, but one can easily show that this algorithm will yield the string $\hat{b} = 111$ with high probability, and hence the algorithm will incorrectly identify them as being far apart. The reason this counterexample works is that postselecting the distributions on seeing the same outcome $y_1 = y_2 = y_3$ heavily skews the distributions towards more likely y_i outputs, and in this example we will almost always have $y = x$, and hence will almost always output $\hat{b} = 111$.

We'll now show that as $n \to \infty$, the ratio of the probabilities between each string tends to 1. Therefore for sufficiently large n, the strings $\hat{b}$ can be make arbitrarily close to equiprobable, so the algorithm works. We'll break into three cases, showing that the strings $\hat{b} = 111$ and 000, 100, and 110 become equiprobable as $n \to \infty$. Since the probability of obtaining a string $\hat{b}$ is only a function of its hamming weight, this will imply all eight possible outcomes for $\hat{b}$ become equiprobable for large n, and hence the error probabilty of the algorithm approaches 1/4 as $n \to \infty$.

Case 1: 111 and 000

Let's consider the extremal case, where $\hat{b} = 111$ or $\hat{b} = 000$. Let $\delta_y = |D_1(y) - D_0(y)|$, so $\sum_y \delta_y \leq \varepsilon$, and furthermore that $D_0(y) \leq D_1(y) + \delta_y$ and $D_1(y) \leq D_0(y) + \delta_y$. Therefore we have that

$$\frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 000|S]} = \frac{\sum_{y\in\{0,1\}^n} D_1(y)^3}{\sum_{y\in\{0,1\}^n} D_0(y)^3} \tag{8.6}$$

$$\leq \frac{\sum_{y\in\{0,1\}^n} (D_0(y) + \delta_y)^3}{\sum_{y\in\{0,1\}^n} D_0(y)^3} \tag{8.7}$$

$$= \frac{\sum_{y\in\{0,1\}^n} D_0(y)^3 + 3D_0(y)^2\delta_y + 3D_0(y)\delta_y^2 + \delta_y^3}{\sum_{y\in\{0,1\}^n} D_0(y)^3} \tag{8.8}$$

$$= 1 + 3\frac{\langle \delta, D_0^2\rangle}{\langle D_0, D_0^2\rangle} + 3\frac{\langle \delta^2, D_0\rangle}{\langle D_0^2, D_0\rangle} + \frac{|\delta^3|_1}{|D_0^3|_1} \tag{8.9}$$

$$\leq 1 + 3\frac{\varepsilon \max_y D_0(y)^2}{\langle D_0, D_0^2\rangle} + 3\frac{\varepsilon^2 \max_y D_0(y)}{\langle D_0^2, D_0\rangle} + \frac{\varepsilon^3}{|D_0^3|_1} \tag{8.10}$$

$$\leq 1 + 3\frac{\varepsilon \max_y D_0(y)^2}{\langle D_0, D_0^2\rangle} + 3\frac{\varepsilon^2 \max_y D_0(y)}{\langle D_0^2, D_0\rangle} + \frac{2^{-3cn}}{2^{-2n}} \tag{8.11}$$

where on line 8.9 we expressed these sums as inner products, on line 8.10 we used the fact the sums in the denominators are maximized when the weight of δ is placed on a single item, line 8.11 follows from the fact the denominator is minimized by the unform distribution. We now need to bound the terms $\frac{\max_y D_0(y)^2}{\langle D_0, D_0^2\rangle}$ and $\frac{\max_y D_0(y)}{\langle D_0, D_0^2\rangle}$ as a function of the universe size $N = 2^n$. One can easily show that the first is upper bounded by $\Theta(N^{2/3})$, and the second is upper bounded by $\Theta(N^{4/3})$.

To see this, let $k = \max_y D_0(y)$, so $2^{-n} \leq k \leq 1$. Then we have that

$$\frac{\max_y D_0(y)^2}{\langle D_0, D_0^2\rangle} \leq \frac{k^2}{k^3 + \frac{(1-k)^3}{(N-1)^2}}$$

because given k, the denominator is minimized by spreading the remaining probability mass evenly over the remaining $N - 1$ elements. By taking the derivative of this as a function of k and setting it equal to zero, we see that the maximum occurs at a solution to the equation $k\left((-5 - (N-1)^2)k^3 + 12k^2 - 9k + 2\right) = 0$. As $N \to \infty$ the real roots of this equation are 0 and $\Theta(N^{-2/3})$ (plus two complex roots), and one can easily show the first is a minimum while the second is the maximum. Hence this quantity is maximized when $k = \Theta(N^{-2/3})$,

which implies the quantity is upper bounded by

$$\frac{\max_y D_0(y)^2}{\langle D_0, D_0^2\rangle} \leq \frac{N^{-4/3}}{N^{-2} + \frac{(1-N^{-2/3})^3}{(N-1)^2}} = \frac{N^{-2/3}}{\Theta(N^{-2})} = \Theta(N^{2/3})$$

A similar proof shows that the second quantity is upper bounded by $\Theta(N^{4/3})$.

Therefore we have that

$$\frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 000|S]} \leq 1 + 3 * 2^{-cn}2^{2n/3} + 3 * 2^{-2cn}2^{4n/3} + \frac{2^{-3cn}}{2^{-2n}} \tag{8.12}$$

$$\leq 1 + o(1) \tag{8.13}$$

since we have $c > 2/3$. Note that the identical proof holds for the case where D_0 and D_1 are switched, therefore we have that $\frac{\Pr[\hat{b} = 000|S]}{\Pr[\hat{b} = 111|S]} \leq 1 + o(1)$ as well. Hence we have

$$1 - o(1) \leq \frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 000|S]} \leq 1 + o(1)$$

So as $n \to \infty$, these strings become equiprobable.

Case 2: 111 and 100 We have that

$$\frac{\Pr[\hat{b} = 111|S]}{\Pr[\hat{b} = 100|S]} = \frac{\sum_{y\in\{0,1\}^n} D_1(y)^3}{\sum_{y\in\{0,1\}^n} D_0(y)^2 D_1(y)} \tag{8.14}$$

$$\leq \frac{\sum_{y\in\{0,1\}^n} D_1(y)(D_0(y)^2 + 2D_0(y)\delta_y + \delta(y)^2)}{\sum_{y\in\{0,1\}^n} D_0(y)^2 D_1(y)} \tag{8.15}$$

$$= 1 + 2\frac{\sum_{y\in\{0,1\}^n} D_1(y)D_0(y)\delta_y}{\sum_{y\in\{0,1\}^n} D_0(y)^2 D_1(y)} + \frac{\sum_{y\in\{0,1\}^n} D_1(y)\delta_y^2}{\sum_{y\in\{0,1\}^n} D_0(y)^2 D_1(y)} \tag{8.16}$$

$$= 1 + 2\frac{\langle \delta_y, D_0 D_1\rangle}{\langle D_0, D_0 D_1\rangle} + \frac{\langle \delta_y^2, D_1\rangle}{\langle D_0^2, D_1\rangle} \tag{8.17}$$

$$\leq 1 + 2\frac{\varepsilon \max_y D_0(y)D_1(y)}{\langle D_0, D_0 D_1\rangle} + \frac{\varepsilon^2 \max_y D_1(y)}{\langle D_0^2, D_1\rangle} \tag{8.18}$$

$$\leq 1 + 2\varepsilon\frac{\max_y D_0(y)^2 + \delta_y D_0(y)}{\langle D_0, D_0 D_1\rangle} + \varepsilon^2\frac{\max_y D_1(y)}{\langle D_0^2, D_1\rangle} \tag{8.19}$$

$$\leq 1 + 2\varepsilon\frac{\max_y D_0(y)^2}{\langle D_0, D_0^2\rangle - \varepsilon \max_y D_0(y)^2} + 2\varepsilon^2\frac{\max_y D_0(y)}{\langle D_0, D_0 D_1\rangle - \varepsilon \max_y D_0(y)^2} \tag{8.20}$$

$$+ \varepsilon^2\frac{\max_y D_1(y)}{\langle D_0, D_0^2\rangle - \varepsilon \max_y D_0(y)^2} \tag{8.21}$$

Where line 8.21 comes from the fact that $D_1(y) \geq D_0(y) - \delta_y$ for all y. We now show that this is upper bounded by $1 + o(1)$, by showing that the term $\frac{\max_y D_0(y)^2}{\langle D_0, D_0^2\rangle - \varepsilon \max_y D_0(y)^2}$, the term $\frac{\max_y D_0(y)}{\langle D_0, D_0 D_1\rangle - \varepsilon \max_y D_0(y)^2}$ and the term $\frac{\max_y D_1(y)}{\langle D_0, D_0^2\rangle - \varepsilon \max_y D_0(y)^2}$ are upper

bounded by $O(2^{2n/3})$, $O(2^{4n/3})$ and $O(2^{4n/3})$, respectively. This, combined with the fact that $\varepsilon = O(2^{-cn})$ for $c > 2/3$, implies that $\frac{\Pr[\hat{b}=111|S]}{\Pr[\hat{b}=100|S]} \leq 1 + o(1)$ as desired.

For the first term, let $k = \max_y D_0(y)$. The this term is upper bounded by

$$\frac{k^2}{k^3 - \frac{k^3}{(N-1)^2} - \varepsilon k^2}$$

because the denominator is minimized by spreading the remaining probability mass evenly over the remaining $N-1$ elements. One can easily show this function is maximized by setting $k = \Theta(N^{-2/3})$. Indeed, taking the derivative of this equation and setting it equal to 0, one can see that the extreme values of k satisfy $k(-2(N-1)^2k^3 - 3k + 2) = 0$. Hence the optimal value of k satisfies $k = \Theta(N^{-2/3})$. For this value of k, the term evalues to $O(N^{2/3})$.

For the second term, if we let k be defined as above, then by the same reasoning we have that the term is uppor bounded by

$$\frac{k}{k^3 - \frac{k^3}{(N-1)^2} - \varepsilon k^2}$$

Again by a similar proof, one can easily show the function is maximized by setting $k = O(N^{-2/3})$, which implies the term is upper bounded by $O(2^{4n/3})$.

For the third term, let $k = \max_y D_1(y)$. By a similar argument as above, we have that

$$\frac{\max_y D_1(y)}{\langle D_0^2, D_1\rangle} \leq \frac{\max_y D_1(y)}{\langle D_1^2, D_1\rangle - 2\langle \delta, D_1^2\rangle + \langle \delta^2, D_1\rangle} \leq \frac{k}{k^3 - \frac{(1-k)^3}{(N-1)^2} - 2\varepsilon k^2 + \varepsilon^2 k}$$

One can show that this term is maximized be setting $k = \Theta(N^{-2/3})$, and therefore this term is upper bounded by $O(N^{4/3})$. Indeed, taking the derivative of this quantity with respect to k and setting it equal to zero, one can see that the maximum value of k satisfies $(-2(N-1)^2 + 2)k^3 + (2\varepsilon - 3)k^2 + 1 = 0$, which implies the maximum value satisfies $k = \Theta(N^{-2/3})$.

We've now shown that $\frac{\Pr[\hat{b}=111|S]}{\Pr[\hat{b}=100|S]} \leq 1 + o(1)$. Now consider the opposite ratio. By the same reasoning as before, we have that

$$\begin{aligned}
\frac{\Pr[\hat{b}=100|S]}{\Pr[\hat{b}=111|S]} &= \frac{\sum_{y\in\{0,1\}^n} D_0(y)^2 D_1(y)}{\sum_{y\in\{0,1\}^n} D_1(y)^3} && (8.22)\\
&\leq 1 + 2\frac{\langle \delta, D_1^2\rangle}{\sum_{y\in\{0,1\}^n} D_1(y)^3} + \frac{\langle \delta^2, D_1\rangle}{\sum_{y\in\{0,1\}^n} D_1(y)^3} && (8.23)\\
&\leq 1 + 2\varepsilon\frac{\max_y D_1(y)}{\sum_{y\in\{0,1\}^n} D_1(y)^3} + \varepsilon^2\frac{\max_y D_1(y)^2}{\sum_{y\in\{0,1\}^n} D_1(y)^3} && (8.24)\\
&\leq 1 + o(1) && (8.25)
\end{aligned}$$

Where on line 8.25 we used the fact that we previously upper bounded these terms when handling Case 1. Hence as $n \to \infty$ the strings $\hat{b} = 111$ and $\hat{b} = 100$ become equiprobable.

Case 3: 111 and 110 We have that

$$\frac{\Pr[\hat{b}=111|S]}{\Pr[\hat{b}=110|S]} = \frac{\sum_{y\in\{0,1\}^n} D_1(y)^3}{\sum_{y\in\{0,1\}^n} D_0(y)D_1(y)^2} \tag{8.26}$$

$$\leq 1 + \frac{\sum_{y\in\{0,1\}^n} \delta_y D_1(y)^2}{\sum_{y\in\{0,1\}^n} D_0(y)D_1(y)^2} \tag{8.27}$$

$$= 1 + \frac{\langle \delta_y, D_1^2\rangle}{\langle D_0, D_1^2\rangle} \tag{8.28}$$

$$\leq 1 + \frac{\varepsilon \max_y D_1(y)^2}{\langle D_0, D_1^2\rangle} \tag{8.29}$$

$$\leq 1 + \frac{\varepsilon \max_y D_1(y)^2}{\langle D_1, D_1^2\rangle - \langle \delta_y, D_1^2\rangle} \tag{8.30}$$

$$\leq 1 + \frac{\varepsilon \max_y D_1(y)^2}{\langle D_1, D_1^2\rangle - \varepsilon \max_y D_1(y)^2} \tag{8.31}$$

$$\leq 1 + o(1) \tag{8.32}$$

Where on line 8.30 we used the fact that $D_0(y) \geq D_1(y) - \delta(y)$, on line 8.31 we used that fact that the numerator is minimized if all the mass of δ_y is placed on the maximum likelihood event of D_1, and on line 8.32 we used the fact that this is the same as the first term we bounded in Case 2.

Now consider the opposite ratio. We have that

$$\frac{\Pr[\hat{b}=110|S]}{\Pr[\hat{b}=111|S]} = \frac{\sum_{y\in\{0,1\}^n} D_0(y)D_1(y)^2}{\sum_{y\in\{0,1\}^n} D_1(y)^3} \tag{8.33}$$

$$\leq \frac{\sum_{y\in\{0,1\}^n} (D_1(y)+\delta(y))D_1(y)^2}{\sum_{y\in\{0,1\}^n} D_1(y)^3} \tag{8.34}$$

$$= 1 + \frac{\sum_{y\in\{0,1\}^n} \delta(y)D_1(y)^2}{\sum_{y\in\{0,1\}^n} D_1(y)^3} \tag{8.35}$$

$$\leq 1 + o(1) \tag{8.36}$$

Where the last line follows from our previous arguments in Case 1. Hence we have that $1 - o(1) \leq \frac{\Pr[\hat{b}=111|S]}{\Pr[\hat{b}=110|S]} \leq 1 + o(1)$ as desired.

Hence we have shown $1 - o(1) \leq \frac{\Pr[\hat{b}=111|S]}{\Pr[\hat{b}=x|S]} \leq 1 + o(1)$ for any three-bit string x. Hence all strings are equiprobable, so in the case the distributions are ε-close, the algorithm's error probability tends to $1/4$ as $n \to \infty$, and hence the algorithm is correct in this case.

To complete the proof, we now show that the probability of error is low then the distributions are $1-\varepsilon$ far apart in total variation distance.

Suppose the distributions are $1-\varepsilon$ far apart in total variation distance. By the definition of total variation distance, there must exist some event $T \subseteq \{0,1\}^n$ for which $|D_0(T) - D_1(T)| \geq 1 - \varepsilon$, where the notation $D_0(T)$ indicates the probability that D_0 outputs an element of the set T, i.e. $D_O(T) = \sum_{y\in T} D_0(y)$. Without loss of generality we have that

$D_0(T) - D_1(T) \geq 1-\varepsilon$, which implies $D_1(\bar{T}) - D_0(\bar{T}) \geq 1-\varepsilon$. Since D_0 and D_1 are probability distributions, this implies $D_0(T) \geq 1-\varepsilon$ and $D_1(T) \leq \varepsilon$, and likewise $D_1(\bar{T}) \geq 1-\varepsilon$ and $D_0(\bar{T}) \leq \varepsilon$. In other words D_0 has almost all its probability mass in T and D_1 has almost all its probability mass in $\bar{T}$.

We'll now show that under these distributions, one will almost certainly see the output $\hat{b} = 000$ or $\hat{b} = 111$. As before, we'll show this by proving that for large n, the strings $\hat{b} = 000$ or 111 are far more likely than $\hat{b} = 001$ or $\hat{b} = 011$, which implies the algorithm almost always outputs the correct answer.

Let $k_0 = \max_{y \in T} D_0(y)$ and let $k_1 = \max_{y \in \bar{T}} D_1(y)$. Suppose without loss of generality that $k_0 \geq k_1$ (otherwise exchange D_0 and D_1 in the argument). Then we have that

$$\frac{\Pr[\hat{b} = 000|S]}{\Pr[\hat{b} = 100|S]} = \frac{\sum_{y \in \{0,1\}^n} D_0(y)^3}{\sum_{y \in \{0,1\}^n} D_0(y)^2 D_1(y)} \tag{8.37}$$

$$= \frac{\langle D_0^2, D_0 \rangle}{\langle D_0^2, D_1 \rangle} \tag{8.38}$$

$$\geq \frac{k_0^3 + \frac{(1-k_0)^3}{(N-1)^2}}{\varepsilon k_0^2 + \varepsilon^2 k_1} \tag{8.39}$$

$$\geq \frac{k_0^3 + \frac{(1-k_0)^3}{(N-1)^2}}{\varepsilon k_0^2 + \varepsilon^2 k_0} \tag{8.40}$$

where line 8.37 follows from the same arguments as the previous section, and line 8.39 follows because the numerator is minimized by placing the uniform distribution on all elements other than the element responsible for k_0, and the denominator is maximized if all the weight that D_0 has on $\bar{T}$ is placed on the element of maximal weight under D_1, and vice versa. Line 8.40 follows from the fact that $k_0 \geq k_1$

Now we show that this quantity is $\omega(1)$, i.e. it approaches infinity as $n \to \infty$. Suppose by contradiction that there exists a constant $c > 1$ which is an upper bound for this quantity. Since $\varepsilon = o(N^{-2/3})$, there exists an n_0 such that for all $n > n_0$, $\varepsilon < \frac{1}{2c} N^{-2/3}$. We claim that for all $n > n_0$, this quantity is greater than c, which is a contradiction.

To see this, we break into three cases.

Case 1: $k_0 \geq N^{-2/3}$

In this case, the numerator is at least k_0^3, while the denominator is at most $\frac{1}{10c} N^{-2/3} k_0^2 + \frac{1}{100c^2} N^{-4/3} k_0 \leq \frac{1}{2c} k_0^3 + \frac{1}{4c^2} k_0^3 < \frac{1}{c} k_0^3$, where the last step follows from the fact that $\frac{1}{2c} \frac{1}{4c^2} < \frac{1}{c}$ for any $c > 1$. Therefore the quantity on line 8.40 is strictly greater than $\frac{k_0^3}{\frac{1}{c} k_0^3} = c$ as desired.

Case 2: $k_0 \leq N^{-2/3}$

In this case the numerator is at least $\frac{(1-k_0)^3}{(N-1)^2}$ which is $\geq 0.75 N^{-2}$ for sufficiently large n, while the denominator is $\varepsilon k_0^2 + \varepsilon^2 k_0 \leq \frac{1}{2c} N^{-2} + \frac{1}{4c^2} N^{-2} < \frac{3}{4c} N^{-2}$ for $c > 1$, which follows from our upper bounds on ε and k_0. Hence the quantity on line 8.40 is strictly greater than c as desired.

Therefore we have shown that and $n \to \infty$, the string $\hat{b} = 000$ (or $\hat{b} = 111$, if $k_0 < k_1$) is much more likely to occur than $\hat{b} = 001$.

A similar proof holds to show that the string $\hat{b} = 000$ (or $\hat{b} = 111$) is more likely to occur than $\hat{b} = 110$; indeed by the same arguments as above, assuming $k_0 \geq k_1$, we have

$$\frac{\Pr[\hat{b} = 000|S]}{\Pr[\hat{b} = 110|S]} \geq \frac{k_0^3 + \frac{(1-k_0)^3}{(N-1)^2}}{\varepsilon^2 k_0 + \varepsilon k_1} \geq \frac{k_0^3 + \frac{(1-k_0)^3}{(N-1)^2}}{\varepsilon^2 k_0 + \varepsilon k_0} \geq \omega(1)$$

Hence the string $\hat{b} = 000$ is far more likely to occur than the strings $\hat{b} = 001$ or $\hat{b} = 011$ (assuming $k_0 > k_1$, otherwise the string $\hat{b} = 111$ is more likely to occur than 001 or 011), and hence the algorithm errs with probability $o(1)$ when the distributions are $1 - \varepsilon$ far apart. This completes the proof.

□

8.6.3 A Weaker Polarization Lower Bound Using Fourier Analysis

Here we show that, if one only cares about black box polarization in the restricted form proposed by Holenstein and Renner [154], then one can prove a lower bound against polarization directly using Fourier analysis alone. This may help the readers understand what's going on in the proof. But please note this result is subsumed by our oracle separation between SZK and PP.

Definition 8.6.3. An (n, ℓ, m)-*special polarizer* is a pair of joint disributions over pairs of strings, (S^0, R^0) and (S^1, R^1), where S^0 and S^1 are over $\{0,1\}^n$, and R^0 and R^1 are over $\{0,1\}^\ell$.

For any distributions D_0 and D_1, we define the polarized distributions $\widehat{D}_0$ and $\widehat{D}_1$ resulting from this polarizer as:

$$\widehat{D}_b = (D_{S_1^b}, \ldots, D_{S_n^b}, R^b)$$

The polarizer then provides the following guarantees:

$$||D_0 - D_1|| > 2/3 \implies ||\widehat{D}_0 - \widehat{D}_1|| > 1 - 2^{-m}$$
$$||D_0 - D_1|| < 1/3 \implies ||\widehat{D}_0 - \widehat{D}_1|| < 2^{-m}$$

An (n, ℓ)-*pseudo polarizer* is the same, except it doesn't provide the above guarantees.

It is to be noted that the technique for polarizing distance between distributions from [211] is a special polarizer. Note also that any (n, ℓ, m)-special polarizer is an (n, ℓ)-pseudo polarizer.

Consider distributions over $\{0,1\}^k$. If there existed a polynomial-time computable (n, ℓ, m)-special polarizer such that $nk + \ell < 2m$, then Theorem 8.6.1 implies that deciding whether pairs of such distributions are close or far can be done in PP. If such a polarizer existed for every k, then this would imply that SZK is contained in PP because of the completeness of the Statistical Distance problem [211]. We rule out this approach of showing such a containment with the following theorem.

Theorem 8.6.4. *For any (n, ℓ, m)-special polarizer, $n = \Omega(m)$.*

Theorem 8.6.4 follows immediately from the following two lemmas. For any $\alpha \in [0,1]$ and bit b, denote by D_b^α the distribution over $\{0,1\}$ that is equal to b with probability $(1+\alpha)/2$. It is easy to see that $||D_0^\alpha - D_1^\alpha|| = \alpha$. We denote by $(\widehat{D}_0^\alpha, \widehat{D}_1^\alpha)$ the distributions that result from applying the special polarizer in the relevant context to (D_0^α, D_1^α) and by $(\widetilde{D}_0^\alpha, \widetilde{D}_1^\alpha)$ the distributions resulting from the pseudo-polarizer.

Lemma 8.6.5. *For any (n,ℓ)-pseudo polarizer and any $\alpha, \beta \in (0,1)$ such that $\alpha > \beta$,*

$$\frac{||\widetilde{D}_0^\alpha - \widetilde{D}_1^\alpha||}{||\widetilde{D}_0^\beta - \widetilde{D}_1^\beta||} \leq 2^{(n+\ell)/2} \left(\frac{\alpha}{\beta}\right)^n$$

Proof. Throughout the proof, we use the symbols for distributions interchangeably with the symbols for vectors representing their mass functions. For each $\alpha \in (0,1)$, we define the following matrix:

$$B_\alpha = \begin{pmatrix} \frac{1+\alpha}{2} & \frac{1-\alpha}{2} \\ \frac{1-\alpha}{2} & \frac{1+\alpha}{2} \end{pmatrix}$$

Consider any distribution p over $\{0,1\}$. The distribution obtained by selecting a bit b according to p and then sampling D_b^α is given by $B_\alpha p$. This can be extended to the case when p is over $\{0,1\}^n$ – if x is drawn according to p, the distribution of $(D_{x_1}^\alpha, \ldots, D_{x_n}^\alpha)$ is given by $B_\alpha^{\otimes n} p$.

Further, if p_0 happens to be the distribution of (S^0, R^0) from an (n,ℓ,m) special polarizer, then $\widetilde{D}_0^\alpha$, when the polarizer is applied to (D_0^α, D_1^α), is given by $(B_\alpha^{\otimes n} \otimes I^{\otimes \ell})p_0$, where I is the 2×2 identity matrix. Similarly, $\widetilde{D}_1^\alpha$ would be $(B_\alpha^{\otimes n} \otimes I^{\otimes \ell})p_1$. Let $C_\alpha = (B_\alpha^{\otimes n} \otimes I^{\otimes \ell})$. We then have:

$$||\widetilde{D}_0^\alpha - \widetilde{D}_1^\alpha|| = \frac{1}{2}||C_\alpha(p_1 - p_0)||_1$$

Both B_α and I have the vectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ as eigenvectors. The corresponding eigenvalues are 1 and α for B_α, and both 1 for I. This implies that the eigenvectors of B are all possible tensor products of these eigenvectors, and the eigenvalue of such a resulting vector is simply the products of the eigenvalues of the vectors that were tensored.

In different terms, the eigenvectors are $(\chi_{T_1} \otimes \chi_{T_2})$ for any $T_1 \subseteq [n]$ and $T_2 \subseteq [\ell]$, which are the characters of $\mathbb{F}_2^{n+\ell}$, and the eigenvalue of this vector would be $\alpha^{|T_1|}$. Since these vectors form a basis, we can write $p_0 = \sum_{T_1,T_2} \widehat{p}_{0,(T_1,T_2)}(\chi_{T_1} \otimes \chi_{T_2})$.

Using the standard relationships between L_1 and L_2 norms, we have the following in-

equalities for any $\alpha, \beta \in (0,1)$ such that $\alpha > \beta$:

$$\begin{aligned}
\frac{||\widehat{D}_0^\alpha - \widehat{D}_1^\alpha||}{||\widehat{D}_0^\beta - \widehat{D}_1^\beta||} &= \frac{\|C_\alpha(p_1 - p_0)\|_1}{\|C_\beta(p_1 - p_0)\|_1} \\
&\leq 2^{(n+\ell)/2} \frac{\|C_\alpha(p_1 - p_0)\|_2}{\|C_\beta(p_1 - p_0)\|_2} \\
&= 2^{(n+\ell)/2} \frac{\|C_\alpha \sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} (\widehat{p}_{1,(T_1,T_2)} - \widehat{p}_{0,(T_1,T_2)})(\chi_{T_1} \otimes \chi_{T_2})\|_2}{\|C_\beta \sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} (\widehat{p}_{1,(T_1,T_2)} - \widehat{p}_{0,(T_1,T_2)})(\chi_{T_1} \otimes \chi_{T_2})\|_2} \\
&= 2^{(n+\ell)/2} \frac{\|\sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} \alpha^{|T_1|} (\widehat{p}_{1,(T_1,T_2)} - \widehat{p}_{0,(T_1,T_2)})(\chi_{T_1} \otimes \chi_{T_2})\|_2}{\|\sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} \beta^{|T_1|} (\widehat{p}_{1,(T_1,T_2)} - \widehat{p}_{0,(T_1,T_2)})(\chi_{T_1} \otimes \chi_{T_2})\|_2} \\
&= 2^{(n+\ell)/2} \left(\frac{\sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} \alpha^{2|T_1|} (\widehat{p}_{1,(T_1,T_2)} - \widehat{p}_{0,(T_1,T_2)})^2}{\sum_{T_1 \subseteq [n], T_2 \subseteq [\ell]} \beta^{2|T_1|} (\widehat{p}_{1,(T_1,T_2)} - \widehat{p}_{0,(T_1,T_2)})^2} \right)^{1/2} \\
&\leq 2^{(n+\ell)/2} \left(\frac{\alpha}{\beta} \right)^n
\end{aligned}$$

where the last inequality follows from the readily verified fact that for any sequences of positive real numbers $\{a_i\}$, $\{b_i\}$, and $\{c_i\}$, $\frac{\sum_i c_i a_i}{\sum_i c_i b_i}$ is at most $\max_i \frac{a_i}{b_i}$. □

Lemma 8.6.6. *For any (n,ℓ)-pseudo polarizer and any $\alpha, \beta \in (0,1)$ such that $\alpha > \beta$, there is an $(n,1)$-pseudo polarizer such that:*

$$\frac{||\widetilde{D}_0^\alpha - \widetilde{D}_1^\alpha||}{||\widetilde{D}_0^\beta - \widetilde{D}_1^\beta||} \geq \frac{||\widehat{D}_0^\alpha - \widehat{D}_1^\alpha||}{||\widehat{D}_0^\beta - \widehat{D}_1^\beta||}$$

Proof. The lemma follows from the following two easily verified facts about Total Variation distance of joint distributions.

Fact 1. *For random variables X, Y and Y',*

$$||(X,Y) - (X,Y')|| = \sum_x \Pr[X = x] ||Y_{|X=x} - Y'_{|X=x}||$$

Fact 2. *For random variables X_0 and X_1 and a uniformly distributed bit B,*

$$||(B, X_B) - (\overline{B}, X_B)|| = ||X_0 - X_1||$$

For convenience, we write the resulting distributions from a polarizer as $\widehat{D}_0^\alpha = (D_{S^0}^\alpha, R^0)$, etc., which is indeed the structure that these distributions have. From the above two facts,

we have the following for a uniformly distributed bit B:

$$\begin{aligned}\frac{||\widehat{D}_0^\alpha - \widehat{D}_1^\alpha||}{||\widehat{D}_0^\beta - \widehat{D}_1^\beta||} &= \frac{||(D_{S^0}^\alpha, R^0) - (D_{S^1}^\alpha, R^1)||}{||(D_{S^0}^\beta, R^0) - (D_{S^1}^\beta, R^1)||} \\ &= \frac{||(B, D_{S^B}^\alpha, R^B) - (\overline{B}, D_{S^B}^\alpha, R^B)||}{||(B, D_{S^B}^\beta, R^B) - (\overline{B}, D_{S^B}^\beta, R^B)||} \\ &= \frac{\sum_r \Pr[R_B = r]||(B, D_{S^B}^\alpha)_{|R^B=r} - (\overline{B}, D_{S^B}^\alpha)_{|R^B=r}||}{\sum_r \Pr[R_B = r]||(B, D_{S^B}^\beta)_{|R^B=r} - (\overline{B}, D_{S^B}^\beta)_{|R^B=r}||} \\ &\leq \max_r \frac{||(B, D_{S^B}^\alpha)_{|R^B=r} - (\overline{B}, D_{S^B}^\alpha)_{|R^B=r}||}{||(B, D_{S^B}^\beta)_{|R^B=r} - (\overline{B}, D_{S^B}^\beta)_{|R^B=r}||}\end{aligned}$$

where the last inequality is from the same argument about sequences of positive numbers as the one at the end of the proof of Lemma 8.6.5.

This proves what we need, as for any r, $((B, D_{S^B})_{|R_B=r}, (\overline{B}, D_{S^B})_{|R_B=r})$ is an $(n, 1)$-pseudo polarizer. □

Proof of Theorem 8.6.4. For any (n, ℓ, m)-special polarizer we have the following when $\alpha = 2/3$ and $\beta = 1/3$:

$$\frac{||\widehat{D}_0^\alpha - \widehat{D}_1^\alpha||}{||\widehat{D}_0^\beta - \widehat{D}_1^\beta||} \geq \frac{1 - 2^{-m}}{2^{-m}} = 2^m - 1$$

Lemmas 8.6.5 and 8.6.6 imply that there is an $(n, 1)$-pseudo polarizer such that:

$$\frac{||\widehat{D}_0^\alpha - \widehat{D}_1^\alpha||}{||\widehat{D}_0^\beta - \widehat{D}_1^\beta||} \leq \frac{||\widetilde{D}_0^\alpha - \widetilde{D}_1^\alpha||}{||\widetilde{D}_0^\beta - \widetilde{D}_1^\beta||} \leq 2^{(n+1)/2} \left(\frac{\alpha}{\beta}\right)^n = 2^{(3n+1)/2}$$

The above two inequalities tell us that $n = \Omega(m)$. □

8.7 Additional Consequences for Property Testing

Lower Bounds for Property Testers That Barely Do Better Than Random Guessing. For any NISZK-hard property testing problem P, our query complexity lower bounds immediately imply that any property testing algorithm for P that outputs the correct answer with probability strictly greater than $1/2$ requires $n^{\Omega(1)}$ queries. For concreteness, we highlight the result we obtain for the NISZK-complete problem of entropy approximation. Specifically, given a distribution D over n elements, a natural problem is to ask how many samples from D are required to estimate the entropy of D to additive error. In 2011, Valiant and Valiant [234] showed that to achieve any constant additive error less than $\log 2/2$, it is both necessary and sufficient to take $\Theta(n/\log n)$ samples from D. However, their bounds assume that one wishes to estimate the entropy with high probability, say with probability $1 - o(1/\text{poly}(n))$. Quantitatively, our $\mathsf{UPP}^{\mathsf{dt}}$ query lower bounds imply the following.

Corollary 8.7.1. *Any algorithm which decides if the entropy of D (over domain size n) is $\leq k - 1$ or $\geq k + 1$ and succeeds with probability $> \frac{1}{2}$ requires $\Omega(n^{1/4}/\log n)$ samples from D.*

In other words, estimating the entropy of a distribution to additive error 2 requires $\tilde{\Omega}(n^{1/4})$ samples , even if the algorithm is only required to have an arbitrarily small bias in deciding the answer correctly.

8.8 Open Problems

Our work leaves a number of open related problems. First, we have shown that the function $\mathsf{GapMaj}(f)$ is hard for $\mathsf{UPP}^{\mathsf{dt}}$, for any function f of high approximate degree, and that $\mathsf{GapAND}(f)$ is hard for $\mathsf{UPP}^{\mathsf{dt}}$, for any function of high positive one-sided approximate degree. Can one extend this work to characterize when $f \circ g$ is hard for $\mathsf{UPP}^{\mathsf{dt}}$, based on some properties of f and g? We conjecture that the $\mathsf{UPP}^{\mathsf{dt}}$ complexity of $\mathsf{GapMaj}(f)$ (respectively, $\mathsf{GapAND}(f)$) is characterized by the *rational approximate degree* of f (respectively, positive one-sided approximate degree of f). Such a result would complement the characterization of the threshold degree of $\mathsf{AND}(f)$ in terms of positive one-sided rational approximate degree given in [218].

Additionally, we have shown a lower bound on certain parameters of the polarization lemma. Is there a polarization algorithm which matches our lower bound?

It would also be interesting to determine whether our lower bounds on property testing algorithms that output the correct answer with probability strictly greater than $1/2$ are quantitatively tight. For example, is there an algorithm that, given query access to a distribution D (over domain size n) that is promised to have entropy $\leq k-1$ or $\geq k+1$, decides which is the case with probability greater than $1/2$, using $\tilde{O}(n^{1/4})$ samples from D?

Finally, the main open question highlighted by our work is to break through the UPP frontier in communication complexity. We formalize this question via the following challenge: prove any superlogarithmic lower bound for an explicit problem in a natural communication model that cannot be efficiently simulated by $\mathsf{UPP}^{\mathsf{cc}}$. Our work shows that any communication model capable of efficiently computing the pattern matrix of $\mathsf{GapMaj}(\mathsf{PTP})$ is a candidate for achieving this goal. Thomas Watson has suggested the following as perhaps the simplest such candidate: consider the $\mathsf{NISZK}^{\mathsf{cc}}$ model, but restricted to be one-way, in the sense that neither Merlin nor Bob can talk to Alice. This model effectively combines the key features of the $\mathsf{NISZK}^{\mathsf{cc}}$ and $\mathbf{OIP}_{+}^{[2]}$ (cf. [86]) communication models. There is a logarithmic cost "one-way NISZK" protocol for the pattern matrix of $\mathsf{GapMaj}(\mathsf{PTP})$, so this model cannot be efficiently simulated by $\mathsf{UPP}^{\mathsf{cc}}$. Curiously, despite the ability of this model to compute functions outside of $\mathsf{UPP}^{\mathsf{cc}}$, to the best of our knowledge it is possible that even the INDEX function requires polynomial cost in this model. Note that while Chakrabarti et al. [86] gave an efficient $\mathbf{OIP}_{+}^{[2]}$ communication protocol for INDEX, their protocol is not zero-knowledge.

Chapter 9

Grover Search and the No-Signaling Principle

In this chapter, we continue exploring the space above BQP by studying modifications of quantum theory which are computationally powerful. Two of the key properties of quantum physics are the no-signaling principle and the Grover search lower bound. That is, despite admitting stronger-than-classical correlations, quantum mechanics does not imply superluminal signaling, and despite a form of exponential parallelism, quantum mechanics does not imply polynomial-time brute force solution of NP-complete problems. Here, we investigate the degree to which these two properties are connected. We examine four classes of deviations from quantum mechanics, for which we draw inspiration from the literature on the black hole information paradox. We show that in these models, the physical resources required to send a superluminal signal scale polynomially with the resources needed to speed up Grover's algorithm. Hence the no-signaling principle is equivalent to the inability to solve NP-hard problems efficiently by brute force within the classes of theories analyzed.

This chapter is based on joint work with Ning Bao and Stephen Jordan [46].

9.1 Introduction

Recently the firewalls paradox [35, 68] has shown that our understanding of quantum mechanics and general relativity appear to be inconsistent at the event horizon of a black hole. (See Section 2.4.2 for a brief overview.) Many of the leading proposals to resolve the paradox involve modifying quantum mechanics. For example, the final-state projection model of Horowitz and Maldecena [155] and the state dependence model of Papadodimas and Raju [203] are modifications to quantum theory which might resolve the inconsistency.

One reason to be skeptical of such modifications of quantum mechanics is that they can often give rise to superluminal signals, and hence introduce acausality into the model. For example, Weinberg nonlinearities allow for superluminal signaling [124, 206]. This is generally seen as unphysical. In contrast, in standard quantum theory, entanglement does not give rise to superluminal signaling.

Another startling feature of such models is that they might allow one to construct computers far more powerful even than conventional quantum computers. In particular, they may allow one to solve NP-hard problems in polynomial time. It is impossible for standard quantum computers to solve NP-hard problems efficiently by searching over all possible solutions. As discussed in section 2.1.3, this is a consequence of the query complexity lower

bound of Bennett, Bernstein, Brassard and Vazirani [56], which shows one cannot search an unstructured list of 2^n items in fewer than $2^{n/2}$ queries with a quantum computer. This bound is achieved by Grover's search algorithm [140]. In contrast, many modifications of quantum theory allow quantum computers to search an exponentially large solution space in polynomial time. For example, quantum computers equipped with postselection [6], Deutschian closed timelike curves [78, 20, 59], or nonlinearities [22, 190, 191, 189, 91] all admit poly-time solution of NP-hard problems by brute force search.

In this chapter we explore the degree to which superluminal signaling and speedups over Grover's algorithm are connected. We consider several modifications of quantum mechanics which are inspired by resolutions of the firewalls paradox. For each modification, we show that the theory admits superluminal signaling if and only if it admits a query complexity speedup over Grover search. Furthermore, we establish a *quantitative* relationship between superluminal signaling and speedups over Grover's algorithm. More precisely, we show that if one can transmit one classical bit of information superluminally using n qubits and m operations, then one can speed up Grover search on a system of $\text{poly}(n, m)$ qubits with $\text{poly}(n, m)$ operations, and vice versa. In other words, the ability to send a superluminal signal with a reasonable amount of physical resources is equivalent to the ability to violate the Grover lower bound with a reasonable amount of physical resources. Therefore the no-signaling principle is equivalent to the inability to solve NP-hard problems efficiently by brute force within the classes of theories analyzed.

Note that in the presence of nonlinear dynamics, density matrices are no longer equivalent to ensembles of pure states. Here, we consider measurements to produce probabilistic ensembles of post-measurement pure states and compute the dynamics of each of these pure states separately. Alternative formulations, in particular Everettian treatment of measurements as entangling unitaries, lead in some cases to different conclusions about superluminal signaling. See e.g. [101].

9.2 Results

We consider four modifications of quantum mechanics, which are inspired by resolutions of the firewalls paradox. The first two are "continuous" modifications in the sense that they have a tunable parameter δ which quantifies the deviation from quantum mechanics. The second two are "discrete" modifications in which standard quantum mechanics is supplemented by one additional operation.

9.2.1 Final state projection

The first "continuous" modification of quantum theory we consider is the final state projection model of Horowitz and Maldecena [155], in which the black hole singularity projects the wavefunction onto a specific quantum state. This can be thought of as a projective measurement with postselection, which induces a linear (but not necessarily unitary) map on the projective Hilbert space. (In some cases it is possible for the Horowitz-Maldecena final state projection model to induce a perfectly unitary process S for the black hole, but in general interactions between the collapsing body and infalling Hawking radiation inside the event horizon induce deviations from unitarity [136].) Such linear but non-unitary maps allow both superluminal signaling and speedups over Grover search. Any non-unitary map M of condition number $1 + \delta$ allows for superluminal signaling with channel capacity $O(\delta^2)$ with a single application of M. The protocol for signaling is simple - suppose Alice has the

ability to apply M, and suppose Alice and Bob share the entangled state

$$\frac{1}{\sqrt{2}}\left(|\phi_0\rangle|0\rangle + |\phi_1\rangle|1\rangle\right), \tag{9.1}$$

where $|\phi_0\rangle$ and $|\phi_1\rangle$ are the minimum/maximum singular vectors of M, respectively. If Alice chooses to apply M or not, then Bob will see a change in his half of the state, which allows signaling with channel capacity $\sim \delta^2$. Furthermore, it is also possible for Bob to signal superluminally to Alice with the same state - if Bob chooses to measure or not to measure his half of the state, it will also affect the state of Alice's system after Alice applies M. So this signaling is bidirectional, even if only one party has access to the non-unitary map. In the context of the black hole information paradox, this implies the acausality in the final state projection model could be present even far away from the black hole. Also, assuming one can apply the same M multiple times, one can perform single-query Grover search using $\sim 1/\delta$ applications of M using the methods of [6, 22]. More detailed proofs of these results are provided in Section 9.4.

We next examine the way in which these results are connected. First, assuming one can speed up Grover search, by a generalization of the hybrid argument of [56], there is a lower bound on the deviation from unitarity required to achieve the speedup. By our previous results this implies a lower bound on the superluminal signaling capacity of the map M. More specifically, suppose that one can search an unstructured list of N items using q queries, with possibly non-unitary operations applied between queries. Then, the same non-unitary dynamics must be capable of transmitting superluminal signals with channel capacity C using shared entangled states, where

$$C = \Omega\left(\left(\frac{\eta}{2q^2} - \frac{2}{N}\right)^2\right) \tag{9.2}$$

Here η is a constant which is roughly ~ 0.42. In particular, solving NP-hard problems in polynomial time by unstructured search would imply superluminal signaling with inverse polynomial channel capacity. This can be regarded as evidence against the possibility of using black hole dynamics to efficiently solve NP-hard problems of reasonable size. A proof of this fact is provided in Section 9.4.

In the other direction, assuming one can send a superluminal signal with channel capacity C, there is a lower bound on the deviation from unitarity which was applied. The proof is provided in Section 9.4. Again by our previous result, this implies one could solve the Grover search problem on a database of size N using a single query and

$$O\left(\frac{\log(N)}{\log(1+C^2)}\right) \tag{9.3}$$

applications of the nonlinear map. Combining these results, this implies that if one can send a superluminal signal with n applications of M, then one can beat Grover's algorithm with $O(n)$ applications of M as well, and vice versa. This shows that in these models, the resources required to observe an exponential speedup over Grover search is polynomially related to the resources needed to send a superluminal signal. Hence an operational version of the no-signaling principle (such as "one cannot observe superluminal signaling in reasonable-sized experiments") is equivalent to an operational version of the Grover lower bound ("one cannot observe violations of the Grover lower bound in reasonable-sized experiments").

9.2.2 Modification of the Born Rule

The next continuous modification of quantum mechanics we consider is modification of the Born rule. Suppose that quantum states evolve by unitary transformations, but upon measurement one sees outcome x with probability proportional to some function $f(\alpha_x)$ of the amplitude α_x on x. That is, one sees x with probability

$$\frac{f(\alpha_x)}{\sum_y f(\alpha_y)} \tag{9.4}$$

Note we have added a normalization factor to ensure this induces a valid probability distribution on outcomes. This is loosely inspired by Marolf and Polchinski's work [186] which suggests that the "state-dependence" resolution of the firewalls paradox [203] gives rise to violations of the Born rule. First, assuming some reasonable conditions on f (namely, that f is differentiable, f' changes signs a finite number of times in $[0, 1]$, and the measurement statistics of f do not depend on the normalization of the state), we must have $f(\alpha_x) = |\alpha_x|^p$ for some p. The proof is provided in Section 9.5.

Next we study the impact of such modified Born rules with $p = 2+\delta$ for small δ. Aaronson [6] previously showed that such models allow for single-query Grover search in polynomial time while incurring a multiplicative overhead $1/|\delta|$, and also allow for superluminal signaling using shared entangled states of $\sim 1/|\delta|$ qubits. (His result further generalizes to the harder problem of *counting* the number of solutions to an NP-hard problem, which is a #P-hard problem). We find that these relationships hold in the opposite directions as well. Specifically, we show if one can send a superluminal signal with an entangled state on m qubits with probability ε, then we must have $\delta = \Omega(\varepsilon/m)$. By the results of Aaronson [6] this implies one can search a list of N items using $O(\frac{m}{\varepsilon} \log N)$ time. Hence having the ability to send a superluminal signal using m qubits implies the ability to perform an exponential speedup of Grover's algorithm with multiplicative overhead m.

In the other direction, if one can achieve even a constant-factor speedup over Grover's algorithm using a system of m qubits, we show $|\delta|$ is at least $1/m$ as well. More precisely, by a generalization of the hybrid argument of [56], if there is an algorithm to search an unordered list of N items with Q queries using m qubits, then

$$\frac{1}{6} \leq \frac{2Q}{\sqrt{N}} + |\delta| \log(M) + O(\delta^2). \tag{9.5}$$

So if $Q < \sqrt{N}/24$, then we must have $|\delta| \geq \frac{1}{12m}$. The proofs of these facts are provided in Section 9.5.

Combining these results shows that the number of qubits required to observe superluminal signaling or even a modest speedup over Grover's algorithm are polynomially related. Hence one can derive an operational version of the no-signaling principle from the Grover lower bound and vice versa. This quantitative result is in some sense stronger than the result we achieve for the final-state projection model, because here we require only a mild speedup over Grover search to derive superluminal signaling.

9.2.3 Cloning, Postselection, and Generic Nonlinearities

We next consider two "discrete" modifications of quantum mechanics in which standard quantum mechanics is supplemented by one additional operation. We show that both modi-

fications admit both superluminal signaling with O(1) qubits and exponential speedups over Grover search.

First, we consider a model in which one can clone single qubits. This model can be easily seen to admit superluminal signaling using entangled states, as pointed out by Aaronson, Bouland, Fitzsimons and Lee [13]. Indeed, suppose two parties Alice and Bob share the state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. If Alice measures her half of the state, and Bob clones his state k times and measures each copy in the computational basis, then Bob will either see either 0^k or 1^k as his output. On the other hand, if Alice does not measure her half of the state, and Bob does the same experiment, his outcomes will be a random string in $\{0,1\}^k$. Bob can distinguish these two cases with an error probability which scales inverse exponentially with k, and thus receive a signal faster than light. In addition to admitting superluminal signaling with entangled states, this model also allows the solution of NP-hard problems (and even #P-hard problems) using a single query to the oracle. This follows by considering the following gadget: given a state ρ on a single qubit, suppose one makes two copies of ρ, performs a Controlled-NOT gate between the copies, and discards one of the copies. This is summarized in a circuit diagram in Fig. 9-1.

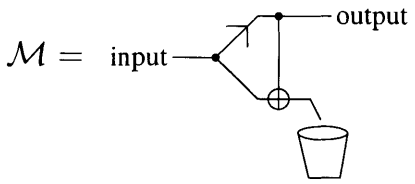


Figure 9-1: Gadget used to show that cloning allows the poly-time solution of NP-hard problems.

This performs a non-linear operation $\mathcal{M}$ on the space of density matrices, and following the techniques of Abrams and Lloyd [22], one can use this operation to "pry apart" quantum states which are exponentially close using polynomially many applications of the gadget. The proof is provided in Section 9.6. This answers our problem posed in Chapter 7 [13] about the power of quantum computers that can clone. Therefore, adding cloning to quantum mechanics allows for both the poly-time solution of NP-hard problems by brute force search, and the ability to efficiently send superluminal signals.

Second, inspired by the final state projection model [155], we consider a model in which one can postselect on a generic state $|\psi\rangle$ of n qubits. Although Aaronson [6] previously showed that allowing for postselection on a single qubit suffices to solve NP-hard and #P-hard problems using a single oracle query, this does not immediately imply that postselecting on a larger state has the same property, because performing the unitary which rotates $|0\rangle^n$ to $|\psi\rangle$ will in general require exponentially many gates. Despite this limitation, this model indeed allows the polynomial-time solution of NP-hard problems (as well as #P-hard problems) and superluminal signaling. To see this, first note that given a gadget to postselect on $|\psi\rangle$, one can obtain multiple copies of $|\psi\rangle$ by inputting the maximally entangled state $\sum_i |i\rangle|i\rangle$ into the circuit and postselecting one register on the state $|\psi\rangle$. So consider creating two copies of $|\psi\rangle$, and applying the gadget shown in Figure 9-2, where the bottom register is postselected onto $|\psi\rangle$, an operation we denote by —$\boxed{|\psi\rangle}$—. For Haar-random $|\psi\rangle$, one can show the quantity $\langle\psi|Z \otimes I|\psi\rangle$ is exponentially small, so this gadget simulates postselection

on $|0\rangle$ on the first qubit. The complete proof is provided in Section 9.7. Therefore, allowing postselection onto generic states is at least as powerful as allowing postselection onto the state $|0\rangle$, so by Aaronson's results [6] this model admits both superluminal signaling and exponential speedups over Grover search.

In addition, we address an open question from [22] regarding the computational implications of general nonlinear maps on pure states. In [22], Abrams and Lloyd argued that generic nonlinear maps allow for the solution of NP-hard problems and #P-hard problems in polynomial time, except possibly for pathological examples. In Section 9.8, we prove this result rigorously in the case the map is differentiable. Thus any pathological examples, if they exist, must fail to be differentiable. (Here we assume the nonlinearity maps pure states to pure states; as a result it does not subsume our results on quantum computers which can clone, as the cloning operation may map pure states to mixed states. A detailed discussion is provided in Section 9.6.) Unfortunately, the action of general nonlinear maps on subsystems of entangled states is not well-defined, essentially because they interact poorly with the linearity of the tensor product. We discuss this in detail in Section 9.9. Hence we are unable to connect this result to signaling in the general case.

9.3 Discussion

The central question in complexity theory is which computational problems can be solved efficiently and which cannot. Through experience, computer scientists have found that the most fruitful way to formalize the notion of efficiency is by demanding that the resources, such as time and memory, used to solve a problem must scale at most polynomially with the size of the problem instance (i.e. the size of the input in bits). A widely held conjecture, called the quantum Church-Turing thesis, states that the set of computational problems solvable in-principle with polynomial resources in our universe is equal to BQP, defined mathematically as the set of decision problems answerable using quantum circuits of polynomially many gates [165]. So far, this conjecture has held up remarkably well. Physical processes which conceivably might be more computationally powerful than quantum Turing machines, such as various quantum many-body dynamics of fermions, bosons, and anyons, as well as scattering processes in relativistic quantum field theories, can all be simulated with polynomial overhead by quantum circuits [11, 121, 160, 162, 161].

The strongest challenge[1] to the quantum Church-Turing thesis comes from quantum gravity. Indeed, many of the recent quantum gravity models proposed in relation to the black hole firewalls paradox involve nonlinear behavior of wavefunctions [155, 203] and thus appear to suggest computational power beyond that of polynomial-size quantum circuits. In

[1] Or rather the only remaining challenge!

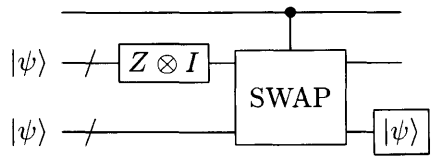


Figure 9-2: Gadget showing postselection onto generic $|\psi\rangle$ is equivalent to postselection onto $|0\rangle$.

particular, the prior work of Abrams and Lloyd suggest that such nonlinearities generically enable polynomial-time solution to NP-hard problems, a dramatic possibility, that standard quantum circuits are not generally expected to admit [22, 3]. Here, we have investigated several models and found a remarkably consistent pattern; in each case, if the modification to quantum mechanics is in a parameter regime allowing polynomial-time solution to NP-hard problems through brute-force search, then it also allows the transmission of superluminal signals through entangled states. Such signaling allows causality to be broken at locations arbitrarily far removed from the vicinity of the black hole, thereby raising serious questions as to the consistency of the models. Thus, the quantum Church-Turing thesis appears to be remarkably robust, not depending in a sensitive way on the complete Hilbert-space formalism of quantum mechanics, but rather derivable from more foundational operational principles such as the impossibility of superluminal signaling. Some more concrete conjectures on these lines are discussed in Section 9.10.

9.4 Proofs: Final-State Projection

Recent developments, particularly the AMPS firewall argument [35], have generated renewed interest in models of black hole physics in which quantum mechanics is modified. Here, we explore some difficulties associated one such scheme, namely the Horowitz-Maldecena final state projection model [155]. In this model, black hole singularities are thought of as boundaries to spacetime with associated boundary conditions on the quantum wavefunction [155]. That is, at the singularity, the wavefunction becomes projected onto a specific quantum state. (This can be thought of as a projective measurement with postselection.)

If one prepares infalling matter in a chosen initial quantum state $|\psi\rangle \in V$, allows it to collapse into a black hole, and then collects all of the the Hawking radiation during the black hole evaporation, one is left with a new quantum state related to the original by some map $S : V \to V$. (We assume that black holes do not alter the dimension of the Hilbert space. Standard quantum mechanics and the Horowitz-Maldecena proposal share this feature.) Within standard quantum mechanics, all such S correspond to multiplication by a unitary matrix, and hence the term S-matrix is used. If one instead drops matter into an existing black hole and collects part of the outgoing Hawking radiation, one is considering an open quantum system. We leave the analysis of this more general scenario to future work.

It is possible for the Horowitz-Maldecena final state projection model to induce a perfectly unitary process S for the black hole. However, as pointed out in [136], interactions between the collapsing body and infalling Hawking radiation inside the event horizon generically induce deviations from unitarity. In this case, the action S of the black hole is obtained by applying some linear but not unitary map M, and then readjusting the norm of the quantum state back to one[2]. Correspondingly, if a subsystem of an entangled state is collapsed into a black hole and the Hawking radiation is collected then the corresponding transformation is $M \otimes I$ followed by an adjustment of the normalization back to 1. Thus, aside from its interest as a potential model for black holes, the Horowitz-Maldecena model provides an interesting example of nonlinear quantum mechanics in which subsystem structure remains well-defined (i.e. the issues described in Section 9.9 do not arise).

In sections 9.4.1 and 9.4.2 we show that if Alice has access to such a black hole and has

[2]Some interpret the final state projection model as inducing a unitary map for observers who stay outside the event horizon, while inducing a non-unitary map for infalling observers [146]. Under this interpretation, our arguments still apply to an infalling observer in the context of a large black hole.

foresightfully shared entangled states with Bob, then Alice can send instantaneous noisy signals to Bob and vice-versa independent of their spatial separation. We quantify the classical information-carrying capacity of the communication channels between Alice and Bob and find that they vanish only quadratically with the deviation from unitarity of the black hole dynamics, as measured by the deviation of the condition number of M from one. Hence, unless the deviation from unitarity is negligibly small, detectable causality violations can infect the entirety of spacetime. Furthermore, the bidirectional nature of the communication makes it possible in principle for Alice to send signals into her own past lightcone, thereby generating grandfather paradoxes.

In section 9.4.3 we consider the use of the black hole dynamical map S to speed up Grover's search algorithm [140]. We find a lower bound on the condition number of M as a function of the beyond-Grover speedup. By our results of sections 9.4.1 and 9.4.2 this in turn implies a lower bound on the superluminal signaling capacity induced by the black hole. In section 9.4.4 we prove the other direction: assuming one can signal superluminally we derive a lower bound on the condition number of M, which in turn implies a super-Grover speedup[3]. We find that the black-box solution of NP-hard problems in polynomial time implies superluminal signaling with inverse polynomial capacity and vice versa.

9.4.1 Communication from Alice to Bob

Theorem 9.4.1. *Suppose Alice and Bob share an entangled state, Alice has access to a black hole described by the Horowitz-Maldecena final state projection model, and Bob is far from the black hole (so is spacelike separated from Alice). Let M be the linear but not necessarily unitary map describing the dynamics of the black hole. The non-unitarity of M is quantified by $\delta = 1 - \kappa$, the deviation of its condition number from one. Alice can transmit instantaneous signals to Bob by choosing to drop her half of a shared entangled state into the black hole or not. The capacity of the resulting classical communication channel from Alice to Bob is at least*

$$C \geq \frac{3}{8 \ln 2} \delta^2.$$

Proof. We prove the lower bound on the channel capacity C by exhibiting an explicit protocol realizing it. Suppose the black hole acts on a d-dimensional Hilbert space and correspondingly M is a $d \times d$ matrix. Then, M has a singular-value decomposition given by

$$M = \sum_{i=0}^{d-1} |\psi_i\rangle \lambda_i \langle \phi_i| \tag{9.6}$$

with

$$\langle \psi_i | \psi_j \rangle = \langle \phi_i | \phi_j \rangle = \delta_{ij}. \tag{9.7}$$

and $\lambda_0, \ldots, \lambda_{d-1}$ all real and nonnegative. We can choose our indexing so that λ_0 is the smallest singular value and λ_1 is largest singular value. Now, suppose Alice and Bob share the state

$$\frac{1}{\sqrt{2}} \left(|\phi_0\rangle |0\rangle + |\phi_1\rangle |1\rangle \right). \tag{9.8}$$

Here $|0\rangle$ and $|1\rangle$ refer to Bob's half of the entangled state, which can be taken to be a qubit. If Alice wishes to transmit the message "0" to Bob she does nothing. If she wishes to

[3] By a "super-Grover speedup", we mean an algorithm which searches an unstructured N-element list using fewer queries than Grover's algorithm.

transmit the message "1" to Bob she applies the black hole dynamical map S to her half of the state. In other words, Alice drops her half of the state into the black hole, and waits for the black hole to evaporate. Correspondingly, one applies $M \otimes I$ to the above state, yielding the unnormalized state

$$\frac{\lambda_0}{\sqrt{2}}|\psi_0\rangle|0\rangle + \frac{\lambda_1}{\sqrt{2}}|\psi_1\rangle|1\rangle. \tag{9.9}$$

After normalization, this becomes:

$$\frac{\lambda_0}{\sqrt{\lambda_0^2+\lambda_1^2}}|\psi_0\rangle|0\rangle + \frac{\lambda_1}{\sqrt{\lambda_0^2+\lambda_1^2}}|\psi_1\rangle|1\rangle. \tag{9.10}$$

Thus, recalling Eq. (9.7), Bob's reduced density matrix in this case is

$$\rho_1 = \frac{\lambda_0^2}{\lambda_0^2+\lambda_1^2}|0\rangle\langle 0| + \frac{\lambda_1^2}{\lambda_0^2+\lambda_1^2}|1\rangle\langle 1|, \tag{9.11}$$

whereas in the case that Alice's message was "0" his reduced density matrix is

$$\rho_0 = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|. \tag{9.12}$$

If M is non-unitary then $\lambda_1 \neq \lambda_0$ and thus the trace distance between these density matrices is nonzero. Consequently, ρ_1 is distinguishable from ρ_0 and some fraction of a bit of classical information has been transmitted to Bob.

More quantitatively, one sees that Bob's optimal measurement is in the computational basis, in which case Alice and Bob are communicating over a classical binary asymmetric channel. Specifically, if Alice transmits a 0, the probability of bit-flip error is $\varepsilon_0 = 1/2$ whereas if Alice transmits a 1, the probability of bit-flip error is

$$\varepsilon_1 = \frac{\lambda_0^2}{\lambda_0^2+\lambda_1^2}. \tag{9.13}$$

A standard calculation (see *e.g* [179]) shows that the classical capacity of this channel is

$$C = h\left(\frac{1}{1+z}\right) - \frac{\log_2(z)}{1+z} + \varepsilon_0 \log_2(z) - h(\varepsilon_0), \tag{9.14}$$

where

$$z = 2^{\frac{h(\varepsilon_1)-h(\varepsilon_0)}{1-\varepsilon_1-\varepsilon_0}} \tag{9.15}$$

and h is the binary entropy

$$h(p) = -p\log_2(p) - (1-p)\log_2(1-p). \tag{9.16}$$

Specializing to $\varepsilon_0 = \frac{1}{2}$ simplifies the expression to

$$C = h\left(\frac{1}{1+y}\right) - \frac{\log_2(y)}{1+y} + \frac{1}{2}\log_2(y) - 1 \tag{9.17}$$

where

$$y = 2^{\frac{h(\varepsilon_1)-1}{1/2-\varepsilon_1}}. \tag{9.18}$$

Lastly, we consider the limiting case $\varepsilon_1 = \frac{1}{2} - \Delta$ for $\Delta \ll 1$. In this limit, we get by Taylor expansion that

$$C = \frac{3}{2\ln 2}\Delta^2 + O(\delta^3). \tag{9.19}$$

By Eq. (9.13), $\Delta = \frac{1}{2}(1-\kappa) + O((1-\kappa)^2)$, which completes the proof. □

9.4.2 Communication from Bob to Alice

Theorem 9.4.2. *Suppose Alice has access to a black hole described by the Horowitz Maldecena final state projection model. Let M be the linear but not necessarily unitary map describing the dynamics of the black hole. The non-unitarity of M is quantified by $\delta = 1 - \kappa$, the deviation of its condition number from one. Bob can transmit instantaneous signals to Alice by choosing to measure his half of a shared entangled state or not. The capacity of the resulting classical communication channel from Bob to Alice is at least*

$$C \geq \frac{3}{8\ln 2}\delta^2.$$

Proof. Suppose again that Alice and Bob share the state $\frac{1}{\sqrt{2}}(|\phi_0\rangle|0\rangle + |\phi_1\rangle|1\rangle)$. If Bob wishes to transmit the message "0" he does nothing, whereas if he wishes to transmit the message "1" he measures his half of the entangled state in the $\{|0\rangle, |1\rangle\}$ basis. Then, Alice applies the black hole dynamical map S to her half of the state[4], and then performs a projective measurement in the basis $\{|\psi_1\rangle, \ldots, |\psi_d\rangle\}$. We now show that this procedure transmits a nonzero amount of classical information from Bob to Alice unless $\lambda_0 = \lambda_1$, in which case M is unitary.

In the case that Bob does nothing, the post-black hole state is again

$$\frac{\lambda_0}{\sqrt{\lambda_0^2 + \lambda_1^2}}|\psi_0\rangle|0\rangle + \frac{\lambda_1}{\sqrt{\lambda_0^2 + \lambda_1^2}}|\psi_1\rangle|1\rangle. \tag{9.20}$$

Thus, Alice's post-black-hole reduced density matrix is

$$\frac{\lambda_0^2}{\lambda_0^2 + \lambda_1^2}|\psi_0\rangle\langle\psi_0| + \frac{\lambda_1^2}{\lambda_0^2 + \lambda_1^2}|\psi_1\rangle\langle\psi_1|. \tag{9.21}$$

Alice's measurement will consequently yield the following probability distribution, given that Bob's message was "0":

$$p(0|0) = \frac{\lambda_0^2}{\lambda_0^2 + \lambda_1^2} \tag{9.22}$$

$$p(1|0) = \frac{\lambda_1^2}{\lambda_0^2 + \lambda_1^2}. \tag{9.23}$$

Now, suppose Bob's message is "1". Then, his measurement outcome will be either $|0\rangle$ or $|1\rangle$ with equal probability. We must analyze these cases separately, since the connection between ensembles of quantum states and density matrices is not preserved under nonlinear

[4]That is, Alice drops her half of the shared state into the black hole, and waits for the black hole to evaporate.

transformations[5]. If he gets outcome zero, then Alice holds the pure state $|\phi_0\rangle$, which gets transformed to $|\psi_0\rangle$ by the action of the black hole. If Bob gets outcome one, then Alice holds $|\phi_1\rangle$, which gets transformed to $|\psi_1\rangle$ by the action of the black hole. Hence, Alice's measurement samples from the following distribution given that Bob's message was "1":

$$p(0|1) = 1/2 \tag{9.24}$$
$$p(1|1) = 1/2. \tag{9.25}$$

Hence, the information transmission capacity from Bob to Alice using this protocol is the same as the Alice-to-Bob capacity calculated in section 9.4.1. □

9.4.3 Super-Grover Speedup implies Superluminal Signaling

Theorem 9.4.3. *Suppose one has access to one or more black holes described by the Horowitz-Maldecena final state projection model. If the non-unitary dynamics induced by the black hole(s) allow the solution of a Grover search problem on a database of size N using q queries then the same non-unitary dynamics could be used transmit instantaneous signals by applying them to half of an entangled state. The capacity of the resulting classical communication channel (bits communicated per use of the nonlinear dynamics) is at least*

$$C = \Omega\left(\left(\frac{\eta}{2q^2} - \frac{2}{N}\right)^2\right)$$

in the regime $0 < \frac{\eta}{2q^2} - \frac{2}{N} \ll 1$, *where* $\eta = \left(\sqrt{2} - \sqrt{2-\sqrt{2}}\right)^2 \simeq 0.42$.

Proof. Let V be the set of normalized vectors in the Hilbert space $\mathbb{C}^N$. We will let $S : V \to V$ denote the nonlinear map that a black hole produces by applying the matrix M and then readjusting the norm of the state to one. We will not assume that all black holes are identical, and therefore, each time we interact with a black hole we may have a different map. We denote the transformation induced on the k^{th} interaction by $S_k : V \to V$. We treat S_k as acting on the same state space for all k, but this is not actually a restriction because we can simply take this to be the span of all the Hilbert spaces upon which the different maps act.

Now suppose we wish to use the operations $S_1, S_2, \ldots$ to speed up Grover search. Let $x \in \{0, \ldots, N-1\}$ denote the solution to the search problem on $\{0, \ldots, N-1\}$. The corresponding unitary oracle on $\mathbb{C}^N$ is[6]

$$O_x = I - 2|x\rangle\langle x|. \tag{9.26}$$

The most general algorithm to search for x is of the form

$$S_q O_x \ldots S_2 O_x S_1 O_x |\psi_0\rangle \tag{9.27}$$

followed by a measurement. Here $|\psi_0\rangle$ is any x-independent quantum state on $\mathbb{C}^N$, and S_k is any transformation that can be achieved on $\mathbb{C}^N$ by any sequence of unitary operations and

[5] Elsewhere we have used density matrices, but only after the application of the nonlinear operation.

[6] An alternate definition is to use a bit-flip oracle $U_x|y\rangle|z\rangle = |y\rangle|z \oplus f(y)\rangle$ where $f(y) = 1$ if $y = x$ and $f(y) = 0$ otherwise. This choice is irrelevant since the phase flip oracle can be simulated using a bit-flip oracle if the output register is initialized to $(|0\rangle - |1\rangle)/\sqrt{2}$, and a bit flip oracle can be simulated using a controlled-phase-flip oracle.

interactions with black holes. Note that our formulation is quite general and includes the case that multiple non-unitary interactions are used after a given oracle query, as is done in [22]. Also, for some k, S_k may be purely unitary. For example, one may have access to only a single black hole, and the rest of the iterations of Grover's algorithm must be done in the ordinary unitary manner. If the final measurement on the state described in Eq. (9.27) succeeds in identifying x with high probability for all $x \in \{0, \ldots, N-1\}$ then we say the query complexity of Grover search using the black hole is at most q.

We now adapt the proof of the $\Omega(\sqrt{N})$ quantum query lower bound for Grover search that was given in[7] [56] to show that any improvement in the query complexity for Grover search implies a corresponding lower bound on the ability of S_k for some $k \in \{1, \ldots, q\}$ to "pry apart" quantum states. This then implies a corresponding lower bound on the rate of a superluminal classical information transmission channel implemented using S_k.

The sequence of quantum states obtained in the algorithm Eq. (9.27) is

$$\begin{aligned} |\psi_0^x\rangle &= |\psi_0\rangle \\ |\phi_1^x\rangle &= O_x|\psi_0\rangle \\ |\psi_1^x\rangle &= S_1 O_x|\psi_0\rangle \\ |\phi_2^x\rangle &= O_x S_1 O_x|\psi_0\rangle \\ |\psi_2^x\rangle &= S_2 O_x S_1 O_x|\psi_0\rangle \\ &\vdots \\ |\psi_q^x\rangle &= S_q O_x \ldots S_1 O_x|\psi_0\rangle. \end{aligned} \tag{9.28}$$

Let

$$|\psi_k\rangle = S_k S_{k-1} \ldots S_1|\psi_0\rangle \tag{9.29}$$

$$C_k = \sum_{x=0}^{N-1} \| |\phi_k^x\rangle - |\psi_{k-1}\rangle \|^2 \tag{9.30}$$

$$D_k = \sum_{x=0}^{N-1} \| |\psi_k^x\rangle - |\psi_k\rangle \|^2. \tag{9.31}$$

$|\psi_k\rangle$ can be interpreted as the state which would have been obtained after the k^{th} step of the algorithm with no oracle queries (or of the Grover search problem lacked a solution).

Now, assume that for all $x \in \{0, \ldots, N-1\}$ the search algorithm succeeds after q queries in finding x with probability at least $\frac{1}{2}$. Then,

$$|\langle x|\psi_q^x\rangle|^2 \geq \frac{1}{2} \quad \forall x \in \{0, \ldots, N-1\} \tag{9.32}$$

which implies

$$D_q \geq \eta N, \tag{9.33}$$

with $\eta = (\sqrt{2} - \sqrt{2-\sqrt{2}})^2 \simeq 0.42$, as shown in [56] and discussed in [199]. By Eq. (9.28),

[7] Our notation is based on the exposition of this proof given in [199].

Eq. (9.30), and Eq. (9.31),

$$C_k = \sum_{x=0}^{N-1} \| O_x |\psi_{k-1}^x\rangle - |\psi_{k-1}\rangle \|^2 \tag{9.34}$$

$$\leq D_{k-1} + 4\sqrt{D_{k-1}} + 4, \tag{9.35}$$

where the above inequality is obtained straightforwardly using the triangle and Cauchy-Schwarz inequalities.

Next, let

$$R_k = D_k - C_k. \tag{9.36}$$

Thus, by Eq. (9.28), Eq. (9.30), and Eq. (9.31),

$$R_k = \sum_{x=0}^{N-1} \| S_k |\phi_k^x\rangle - S_k |\psi_{k-1}\rangle \|^2 - \sum_{x=0}^{N-1} \| |\phi_k^x\rangle - |\psi_{k-1}\rangle \|^2. \tag{9.37}$$

Hence, one sees that R_k is some measure of the ability of S_k to "pry apart" quantum states. (In ordinary quantum mechanics S_k would be unitary and hence R_k would equal zero.)

Combining Eq. (9.36) and Eq. (9.35) yields

$$D_k \leq R_k + D_{k-1} + 4\sqrt{D_{k-1}} + 4. \tag{9.38}$$

Let

$$B = \max_{1 \leq k \leq q} R_k. \tag{9.39}$$

Then Eq. (9.38) yields the simpler inequality

$$D_k \leq B + D_{k-1} + 4\sqrt{D_{k-1}} + 4. \tag{9.40}$$

By Eq. (9.28) and Eq. (9.31),

$$D_0 = 0. \tag{9.41}$$

By an inductive argument, one finds that Eq. (9.40) and Eq. (9.41) imply

$$D_k \leq (4 + B)k^2. \tag{9.42}$$

Combining Eq. (9.42) and Eq. (9.33) yields

$$(4 + B)q^2 \geq \eta N, \tag{9.43}$$

or in other words

$$B \geq \frac{\eta N}{q^2} - 4. \tag{9.44}$$

Thus, by Eq. (9.39) and Eq. (9.37), there exists some $k \in \{1, \ldots, q\}$ such that

$$\sum_{x=0}^{N-1} \left(\| S_k |\phi_k^x\rangle - S_k |\psi_{k-1}\rangle \|^2 - \| |\phi_k^x\rangle - |\psi_{k-1}\rangle \|^2 \right) \geq \frac{\eta N}{q^2} - 4 \tag{9.45}$$

Hence, there exists some $x \in \{0, \ldots, N-1\}$ such that

$$\||S_k|\phi_k^x\rangle - S_k|\psi_{k-1}\rangle\|^2 - \||\phi_k^x\rangle - |\psi_{k-1}\rangle\|^2 \geq \frac{\eta}{q^2} - \frac{4}{N}. \tag{9.46}$$

To simplify notation, define

$$\begin{aligned} |A\rangle &= |\phi_k^x\rangle & (9.47)\\ |B\rangle &= |\psi_{k-1}\rangle & (9.48)\\ |A'\rangle &= S_k|\phi_k^x\rangle & (9.49)\\ |B'\rangle &= S_k|\psi_{k-1}\rangle. & (9.50) \end{aligned}$$

Then Eq. (9.46) becomes

$$\||A'\rangle - |B'\rangle\|^2 - \||A\rangle - |B\rangle\|^2 \geq \frac{\eta}{q^2} - \frac{4}{N}. \tag{9.51}$$

Recalling that $\||\psi\rangle\| = \sqrt{\langle\psi|\psi\rangle}$, Eq. (9.51) is equivalent to

$$\mathrm{Re}\langle A|B\rangle - \mathrm{Re}\langle A'|B'\rangle \geq \varepsilon \tag{9.52}$$

with

$$\varepsilon = \frac{\eta}{2q^2} - \frac{2}{N}. \tag{9.53}$$

Next we will show that, within the framework of final-state projection models, Eq. (9.52) implies that Alice can send a polynomial fraction of a bit to Bob or vice versa using preshared entanglement and a single application of black hole dynamics. Recall that, within the final state projection model,

$$\begin{aligned} |A'\rangle &= \frac{M|A\rangle}{\sqrt{\langle A|M^\dagger M|A\rangle}} & (9.54)\\ |B'\rangle &= \frac{M|B\rangle}{\sqrt{\langle B|M^\dagger M|B\rangle}} & (9.55) \end{aligned}$$

Thus, Eq. (9.52) is equivalent to

$$\mathrm{Re}\left[\langle A|\left(I - \frac{M^\dagger M}{\sqrt{\langle A|M^\dagger M|A\rangle\langle B|M^\dagger M|B\rangle}}\right)|B\rangle\right] \geq \varepsilon \tag{9.56}$$

Hence,

$$\left\| I - \frac{M^\dagger M}{\sqrt{\langle A|M^\dagger M|A\rangle\langle B|M^\dagger M|B\rangle}} \right\| \geq \varepsilon. \tag{9.57}$$

Again using λ_0 to denote the smallest singular value of M and λ_1 to denote the largest, we see that, assuming ε is nonnegative, Eq. (9.57) implies either

Case 1:

$$\frac{\lambda_1^2}{\sqrt{\langle A|M^\dagger M|A\rangle\langle B|M^\dagger M|B\rangle}} \geq 1 + \varepsilon, \tag{9.58}$$

which implies

$$\frac{\lambda_1^2}{\lambda_0^2} \geq 1 + \varepsilon, \tag{9.59}$$

or

Case 2:

$$\frac{\lambda_0^2}{\sqrt{\langle A|M^\dagger M|A\rangle\langle B|M^\dagger M|B\rangle}} \leq 1 - \varepsilon, \tag{9.60}$$

which implies

$$\frac{\lambda_0^2}{\lambda_1^2} \leq 1 - \varepsilon. \tag{9.61}$$

Examining Eq. (9.52), one sees that ε can be at most 2. If $0 \leq \varepsilon \leq 1$ then Eq. (9.61) implies Eq. (9.59). If $1 < \varepsilon \leq 2$ then case 2 is impossible. Hence, for any nonnegative ε one obtains Eq. (9.59). Hence, by the results of sections 9.4.1 and 9.4.2, Alice and Bob can communicate in either direction through a binary asymmetric channel whose bitflip probabilities ε_0 for transmission of zero and ε_1 for transmission of one are given by

$$\varepsilon_0 = \frac{1}{2} \tag{9.62}$$

$$\varepsilon_1 = \frac{\lambda_0^2}{\lambda_0^2 + \lambda_1^2} \leq \frac{1}{2+\varepsilon}. \tag{9.63}$$

For $0 \leq \varepsilon \leq 2$, $\frac{1}{2+\varepsilon} \leq \frac{1}{2} - \frac{\varepsilon}{8}$. Thus, Eq. (9.63) implies the following more convenient inequality

$$\varepsilon_1 \leq \frac{1}{2} - \frac{\varepsilon}{8}. \tag{9.64}$$

In section 9.4.1 we calculated that the channel capacity in the case that $\varepsilon_0 = \frac{1}{2}$ and $\varepsilon_1 = \frac{1}{2} - \delta$ is $\Omega(\delta^2)$ for $\delta \ll 1$. Thus, Eq. (9.62) and Eq. (9.64) imply a channel capacity in either direction of

$$C = \Omega\left(\left(\frac{\eta}{2q^2} - \frac{2}{N}\right)^2\right) \tag{9.65}$$

in the regime $0 < \frac{\eta}{2q^2} - \frac{2}{N} \ll 1$. □

The above scaling of the superluminal channel capacity with Grover speedup shows that polynomial speedup for small instances or exponential speedup for large instances imply 1/poly superluminal channel capacity. In particular, to solve NP in polynomial time without exploiting problem structure we would need $q \propto \log^c N$ for some constant c. In this setting $N = 2^n$ where n is the size of the witness for the problem in NP. In this limit, Eq. (9.65) implies instantaneous signaling channels in each direction with capacity at least

$$C = \Omega\left(\frac{1}{\log^{4c} N}\right) = \Omega\left(\frac{1}{n^{4c}}\right). \tag{9.66}$$

If we assume that superluminal signaling capacity is limited to some negligibly small capacity $C \leq \varepsilon$ then, by Eq. (9.66), NP-hard problems cannot be solved by unstructured search in time scaling polynomially with witness size (specifically n^c for some constant c) except

possibly for unphysically large instances with $n = \Omega\left(\left(\frac{1}{\varepsilon}\right)^{\frac{1}{4c}}\right)$.

9.4.4 Signaling implies Super-Grover Speedup

In sections 9.4.1 and 9.4.2 we showed that if final-state projection can be used to speed up Grover search it can also be used for superluminal signaling. In this section we show the converse. Unlike in section 9.4.3, we here make the assumption that we can make multiple uses of the same non-unitary map S (just as other quantum gates can be used multiple times without variation). Since signaling cannot be achieved by performing unitary operations on entangled quantum degrees of freedom, superluminal signaling implies non-unitarity. Furthermore, as shown in Section 9.9, iterated application of any nonlinear but differentiable map allows the Grover search problem to be solved with only a single oracle query. The nonlinear maps that arise in final-state projection models are differentiable (provided M is invertible), and thus within the final-state projection framework signaling implies single-query Grover search. In the remainder of this section we quantitatively investigate how many iterations of the nonlinear map are needed to achieve single-query Grover search, as a function of the superluminal signaling capacity. We find that unless the signaling capacity is exponentially small, logarithmically many iterations suffice. Specifically, our main result of this section is the following theorem.

Theorem 9.4.4. *Suppose Alice has access to a linear but not necessarily unitary maps on quantum states, as can arise in the Horowitz-Maldecena final state projection model. Suppose she achieves instantaneous classical communication capacity of C bits transmitted per use of nonunitary dynamics. Then she could solve the Grover search problem on a database of size N using a single query and $O\left(\frac{\log(N)}{\log(1+C^2)}\right)$ applications of the available nonunitary maps.*

Proof. Suppose Alice has access to black hole(s) and Bob does not. Alice will use this to send signals to Bob using some shared entangled state $|\psi\rangle_{AB}$. Her most general protocol is to apply some map M_0 to her half of the state if she wishes to transmit a zero and some other map M_1 if she wishes to transmit a one. (As a special case, M_0 could be the identity.) Here, per the final state projection model, M_0 and M_1 are linear but not necessarily unitary maps, and normalization of quantum states is to be adjusted back to one after application of these maps. The possible states shared by Alice and Bob given Alice's two possible messages are

$$|\psi_0\rangle_{AB} \propto M_0|\psi\rangle_{AB} \tag{9.67}$$
$$|\psi_1\rangle_{AB} \propto M_1|\psi\rangle_{AB}. \tag{9.68}$$

The signaling capacity is determined by the distinguishability of the two corresponding reduced density matrices held by Bob

$$\rho_0 = \mathrm{Tr}_A\left[|\psi_0\rangle_{AB}\right] \tag{9.69}$$
$$\rho_1 = \mathrm{Tr}_A\left[|\psi_1\rangle_{AB}\right]. \tag{9.70}$$

We can define

$$|\psi'\rangle \propto M_0|\psi\rangle_{AB} \tag{9.71}$$

in which case

$$|\psi_1\rangle_{AB} \propto M_1 M_0^{-1} |\psi'\rangle. \tag{9.72}$$

(We normalize $|\psi'\rangle$ so that $\langle\psi'|\psi'\rangle = 1$.) Thus, the signaling capacity is determined by the distinguishability of

$$\rho_0 = \mathrm{Tr}_A \left[|\psi'\rangle\right] \tag{9.73}$$

$$\rho_1 = \mathrm{Tr}_A \left[\frac{1}{\eta} M |\psi'\rangle\right] \tag{9.74}$$

where

$$M = M_1 M_0^{-1} \tag{9.75}$$

$$\eta = \sqrt{\langle\psi'|M^\dagger M|\psi'\rangle}. \tag{9.76}$$

We have thus reduced our analysis to the case that Alice applies some non-unitary map M to her state if she wants to transmit a one and does nothing if she wants to transmit a zero. We will next obtain a lower bound $\kappa_{\min}$ on the condition number of M as a function of the signaling capacity from Alice to Bob. This then implies that one of M_0, M_1 has a condition number at least $\sqrt{\kappa_{\min}}$ for the general case.

Suppose that M has the following singular value decomposition

$$M = \sum_i \lambda_i |\psi_i\rangle\langle\phi_i|. \tag{9.77}$$

We can express $|\psi'\rangle$ as

$$|\psi'\rangle = \sum_{i,j} \alpha_{ij} |\phi_i\rangle |B_j\rangle \tag{9.78}$$

where $|\phi_1\rangle, |\phi_2\rangle, \ldots$ is the basis determined by the singular value decomposition Eq. (9.77) and $|B_1\rangle, |B_2\rangle, \ldots$ is the basis Bob will perform his measurement in when he tries to extract Alice's message. If Alice wishes to transmit one then she applies M yielding

$$|\psi_1\rangle \propto \sum_{i,j} \lambda_i \alpha_{ij} |\psi_i\rangle |B_j\rangle. \tag{9.79}$$

So

$$\rho_0 = \sum_{i,j,k} \alpha_{ij} \alpha_{ik}^* |B_j\rangle\langle B_k| \tag{9.80}$$

$$\rho_1 = \sum_{i,j,k} \frac{\lambda_i^2}{\eta} \alpha_{ij} \alpha_{ik}^* |B_j\rangle\langle B_k|. \tag{9.81}$$

Consequently, Bob's measurement will yield a sample from the following probability distri-

butions conditioned on Alice's message.

$$p(j|0) = \sum_i |\alpha_{ij}|^2 \tag{9.82}$$

$$p(j|1) = \sum_i \frac{\lambda_i^2}{\eta} |\alpha_{ij}|^2. \tag{9.83}$$

The total variation distance between these distributions, which determines the capacity of the superluminal channel is

$$\Delta = \frac{1}{2} \sum_j |p(j|0) - p(j|1)| = \frac{1}{2} \sum_j \left| \sum_i |\alpha_{ij}|^2 \left(1 - \frac{\lambda_i^2}{\eta}\right) \right|. \tag{9.84}$$

From a given value of this total variation distance we wish to derive a lower bound on the condition number of M, that is, the ratio of the largest singular value to the smallest. Applying the triangle inequality to Eq. (9.84) yields

$$\Delta \leq \frac{1}{2} \sum_{ij} |\alpha_{ij}|^2 \left| 1 - \frac{\lambda_i^2}{\eta} \right|. \tag{9.85}$$

Because α_{ij} are amplitudes in a normalized quantum state,

$$p(i) = \sum_j |\alpha_{ij}|^2 \tag{9.86}$$

is a probability distribution. We can thus rewrite Eq. (9.85) as

$$\Delta \leq \frac{1}{2} \sum_i p(i) \left| 1 - \frac{\lambda_i^2}{\eta} \right| \tag{9.87}$$

$$\leq \frac{1}{2} \max_i \left| 1 - \frac{\lambda_i^2}{\eta} \right|. \tag{9.88}$$

In keeping with the notation from previous sections, we let λ_0 denote the smallest singular value of M and λ_1 the largest. Thus, Eq. (9.88) yields

$$\Delta \leq \frac{1}{2} \max \left\{ 1 - \frac{\lambda_0^2}{\eta}, \frac{\lambda_1^2}{\eta} - 1 \right\}. \tag{9.89}$$

Similarly,

$$\eta = \sum_{jk} |\alpha_{jk}|^2 \lambda_j^2 \tag{9.90}$$

$$= p(j) \lambda_j^2 \tag{9.91}$$

$$\in [\lambda_0^2, \lambda_1^2]. \tag{9.92}$$

Applying Eq. (9.92) to Eq. (9.89) yields

$$\Delta \leq \frac{1}{2} \max \left\{ 1 - \frac{\lambda_0^2}{\lambda_1^2}, \frac{\lambda_1^2}{\lambda_0^2} - 1 \right\}. \tag{9.93}$$

As shown in section 9.4.5, the channel capacity C is related to the total variation distance Δ according to

$$C \leq \Delta - \Delta \log_2 \Delta \tag{9.94}$$

for $\Delta \leq 1/e$. For small Δ, the $-\Delta \log_2 \Delta$ term dominates the Δ term. We can simplify further by noting that for all positive Δ, $\sqrt{\Delta} > -\Delta \log_2(\Delta)$. Hence, $C = O(\sqrt{\Delta})$. Thus to achieve a given channel capacity C we need

$$\Delta = \Omega(C^2). \tag{9.95}$$

By Eq. (9.93), this implies that achieving a channel capacity C requires

$$|1 - \kappa_{\min}^2| = \Omega(C^2), \tag{9.96}$$

where $\kappa_{\min}$ is the condition number of the nonlinear map $M = M_1 M_0^{-1}$. This implies that one of M_0 or M_1 must have condition number at least $\kappa = \sqrt{\kappa_{min}} = \Omega\left((1 - C^2)^{1/4}\right)$. This in turn implies Grover search with one query and $O(\log_\kappa(N))$ applications of the nonlinear map via the methods of [22]. □

9.4.5 Channel Capacity and Total Variation Distance

Alice wishes to transmit a message to Bob. If she sends zero Bob receives a sample from $p(B|0)$ and if she sends one Bob receives a sample from $p(B|1)$. Here, B is a random variable on a finite state space $\Gamma = \{0, 1 \ldots, d-1\}$. The only thing we know about this channel is that

$$|p(B|0) - p(B|1)| = \delta, \tag{9.97}$$

where $|\cdot|$ denotes the total variation distance (*i.e.* half the l_1 distance). In this section we derive an upper bound on the channel capacity as a function of δ. Specifically, we show that the (asymptotic) capacity C obeys

$$C \leq \delta - \delta \log_2 \delta. \tag{9.98}$$

Any strategy that Bob could use for decoding Alice's message corresponds to a decomposition of Γ as

$$\Gamma = \Gamma_0 \sqcup \Gamma_1 \tag{9.99}$$

where Γ_0 are the outcomes that Bob interprets as zero and Γ_1 are the outcomes that Bob interprets as one.

From Eq. (9.97) it follows that

$$|p(b \in \Gamma_0 | A = 0) - p(b \in \Gamma_0 | A = 1)| \leq \delta. \tag{9.100}$$

(The defining property of total variation distance is that this holds for any set Γ_0.)

Let $F = 0$ whenever $B \in \Gamma_0$ and $F = 1$ whenever $B \in \Gamma_1$. That is, the random variable F is Bob's guess as to Alice's message. By standard Shannon theory [95], the channel capacity is the mutual information $I(F; A)$ maximized over Alice's choice of $p(A)$.

From Eq. (9.100) it follows that

$$|p(F|A = 0) - p(F|A = 1)| \leq \delta. \tag{9.101}$$

Let p_α be the probability distribution

$$p_\alpha(F) = \alpha p(F|A=0) + (1-\alpha)p(F|A=1) \tag{9.102}$$

for some $\alpha \in [0,1]$. From the elementary properties of total variation distance it follows that

$$|p_\alpha(F) - p(F|A=0)| \leq \delta \tag{9.103}$$

and

$$|p_\alpha(F) - p(F|A=1)| \leq \delta \tag{9.104}$$

for any choice of α. In particular, we may set $\alpha = p(A=0)$, in which case we have

$$\begin{aligned} |p(F) - p(F|A=0)| &\leq \delta & (9.105)\\ |p(F) - p(F|A=1)| &\leq \delta. & (9.106) \end{aligned}$$

Next, we recall the Fannes inequality. This says that for any two density matrices ρ, σ on a d-dimensional Hilbert space whose trace distance satisfies $T \leq \frac{1}{e}$

$$|S(\rho) - S(\sigma)| \leq T \log_2 d - T \log_2 T. \tag{9.107}$$

Specializing to the special case that σ and ρ are simultaneously diagonalizable, one obtains the following statement about classical entropies.

Corollary 9.4.1. *Let p and q be two probability distributions on a state space of size d. Let T be the total variation distance between p and q. Suppose $T \leq \frac{1}{e}$. Then*

$$|H(p) - H(q)| \leq T \log_2 d - T \log_2 T. \tag{9.108}$$

Applying corollary 9.4.1 to Eq. (9.105) and Eq. (9.106) yields[8]

$$\begin{aligned} |H[p(F)] - H[p(F|A=0)]| &\leq \delta - \delta \log_2 \delta & (9.109)\\ |H[p(F)] - H[p(F|A=1)]| &\leq \delta - \delta \log_2 \delta. & (9.110) \end{aligned}$$

Thus,

$$\begin{aligned} I(F;A) &= H(F) - H(F|A) & (9.111)\\ &= H[p(F)] - p(A=0)H[p(F|A=0)] - p(A=1)H[p(F|A=1)] & (9.112)\\ &\leq \delta - \delta \log_2 \delta, & (9.113) \end{aligned}$$

which completes the derivation.

9.5 Proofs: Violations of the Born Rule

In this Section we consider modification of quantum mechanics in which states evolve unitarily, but measurement statistics are not given by the Born rule. This is loosely inspired by the "state dependence" resolution of the firewalls paradox, put forth by Papadodimas and

[8] We have used $H[p]$ to denote the entropy of a probability distribution p and $H(R)$ to denote the entropy of a random variable R.

Raju [203]. In this theory, the measurement operators O which correspond to observables are not fixed linear operators, but rather vary depending on the state they are operating on, i.e. $O = O(|\psi\rangle)$. (In general such dependencies lead to nonlinearities in quantum mechanics, but Papadodimas and Raju argue these are unobservable in physically reasonable experiments.) Recently Marolf and Polchinski [186] have claimed that such modifications of quantum mechanics lead to violations of the Born rule. We do not take a position either way on Marolf and Polchinski's claim, but use it as a starting point to investigate how violations of the Born rule are related to superluminal signaling and computational complexity.

Here we consider violations of the Born rule of the following form: given a state $|\psi\rangle = \sum_x \alpha_x |x\rangle$, the probability p_x of seeing outcome x is given by

$$p_x = \frac{f(\alpha_x)}{\sum_{x'} f(\alpha_{x'})} \tag{9.114}$$

for some function $f(\alpha) : \mathbb{C} \to \mathbb{R}^+$. We assume that states in the theory evolve unitarily as in standard quantum mechanics. One could consider more general violations of the Born rule, in which the function f depends not only on the amplitude α_x on x, but on the amplitudes on other basis states as well. However such a generalized theory seems impractical to work with, so we do not consider such a theory here.

We first show that, assuming a few reasonable conditions on f (namely that f has a reasonably behaved derivative and that measurement statistics do not depend on the normalization of the state), the only way to modify the Born rule is to set $f(\alpha) = |\alpha|^{2+\delta}$ for some $\delta \neq 0$. We then show that in theories where the Born rule is modified, superluminal signaling is equivalent to a speedup to Grover search. More precisely, we show that if one can send superluminal signals using states on n qubits, then one can speed up Grover search on a system with $O(n)$ qubits, and vice versa. Hence one can observe superluminal signals on reasonably sized systems if and only if one can speed up Grover search using a reasonable number of qubits.

We are not the first authors to examine the complexity theoretic consequences of modifications to the Born rule. Aaronson [6] considered such modifications, and showed that if δ is any constant, then such modifications allow for the solution of #P-hard problems in polynomial time. Our contributions are 1) to show the opposite direction, namely that a significant speedup over Grover search implies the deviation from the Born rule δ, and 2) to connect this to superluminal signaling.

We prove our results in several steps. First, in Theorems 9.5.2 and 9.5.3, we show that deviations in the Born rule by δ allow the solution of NP-hard problems and superluminal signaling using $O(1/\delta)$ qubits. As noted previously, Theorem 9.5.2 follows from the work of Aaronson [6], but we include a proof for completeness.

In Theorem 9.5.4 we show that, assuming one has a superluminal signaling protocol using a shared state on m qubits, the deviation from the Born rule δ must be $\geq \Omega(1/m)$. Likewise in Theorem 9.5.5 we show that if one can achieve a constant factor super-Grover speedup using m qubits, that we must have $\delta \geq \Omega(1/m)$ as well. Combining these with Theorems 9.5.2 and 9.5.3 shows that a super-Grover speedup on m qubits implies superluminal signaling protocols with $O(m)$ qubits and vice versa. Supplementary Figure 1 explains the relationship between these theorems below.

In short, we find that a violation of the Born rule by δ is equivalent to allowing a super-Grover speedup and an instantaneous signaling protocol using $1/\delta$ qubits. Hence in theories

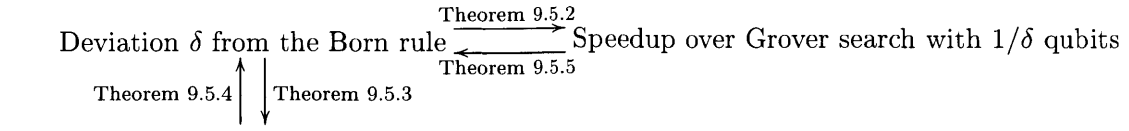


Figure 9-3: Relationship between theorems connecting signaling and search.

in which δ is only polynomially suppressed (as a function e.g. of the number of fields N in Super-Yang-Mills), then such theories allow for superluminal signaling and violations of the Grover lower bound with reasonable overheads. On the other hand, our results do not rule out violations of the Born rule in which $1/\delta$ is unphysically large.

9.5.1 Power law violations are unique

We now show that, given some reasonable assumptions about the function $f(\alpha)$, the only possible violation of the Born rule is given by $f(\alpha) = |\alpha|^p$. In particular we will demand the following properties of f:

1. Well-behaved derivative: $f(\alpha)$ is continuous and differentiable, and $f'(\alpha)$ changes sign at most a finite number of times on $[0, 1]$

2. Scale invariance: for any $k \in \mathbb{C}$, we have that $\frac{f(k\alpha)}{\sum_x f(k\alpha_x)} = \frac{f(\alpha)}{\sum_x f(\alpha_x)}$. In other words the calculation of the probability p_x of seeing outcome x is independent of the norm or phase of the input state; it only depends on the state of the projective Hilbert space.

There are a number of other reasonable constraints one could impose; for instance one could demand that the modified Born rule has to behave well under tensor products. Suppose you have a state $|\psi\rangle = \sum_x \alpha_x |y\rangle$ and a state $|\phi\rangle = \sum_y \beta_y |y\rangle$. A reasonable assumption would be to impose that in the state $|\psi\rangle \otimes |\phi\rangle$, the probability p_{xy} of measuring outcome xy should be equal to $p_x p_y$, i.e. a tensor product state is equivalent to independent copies of each state. More formally this would state that

$$\frac{f(\alpha_x \beta_y)}{\sum_{x'y'} f(\alpha_{x'} \beta_{y'})} = \frac{f(\alpha_x)}{\sum_{x'} f(\alpha_{x'})} \frac{f(\beta_y)}{\sum_{y'} f(\beta_{y'})}. \tag{9.115}$$

Let us call this the Tensor product property. It will turn out that the Tensor product property is implied by the Scale invariance property, which we will show in our proof.

We now show that the Well-behaved derivative and Scale invariance properties imply $f(\alpha) = |\alpha|^p$ for some p.

Theorem 9.5.1. *Suppose that f satisfies the Well-behaved derivative and Scale invariance properties. Then $f(\alpha) = |\alpha|^p$ for some $p \in \mathbb{R}$.*

Proof. First, note that the functions $f(\alpha)$ and $cf(\alpha)$ give the same measurement statistics for any scalar $c \in \mathbb{R}$. To eliminate this redundancy in our description of f, we'll choose c such that $f(1) = 1$.

For any $\alpha \in \mathbb{C}$, consider the (non-normalized) state $\alpha|0\rangle + |1\rangle$. By scale invariance, for any $\beta \in \mathbb{C}$, we must have that

$$\frac{f(\alpha)}{f(\alpha)+f(1)} = \frac{f(\alpha\beta)}{f(\alpha\beta)+f(\beta)} \tag{9.116}$$

which implies that $f(\alpha)f(\beta) = f(\alpha\beta)f(1) = f(\alpha\beta)$ for all $\alpha, \beta \in \mathbb{C}$. One can easily check that this implies the tensor product property.

In particular this holds for any phase, so if $\alpha = |\alpha|e^{i\theta}$, we must have that $f(\alpha) = \hat{f}(|\alpha|)g(\theta)$ for some functions $\hat{f} : \mathbb{R}^{\geq 0} \to \mathbb{R}^+$ and $g : [0, 2\pi) \to \mathbb{R}^+$. Note that taking $g \to cg$ and $\hat{f} \to \hat{f}/c$ leaves f invariant for any scalar $c \in \mathbb{R}^+$. So without loss of generality, since $f(1) = 1$, we can set $\hat{f}(1) = g(0) = 1$ as well by an appropriate choice of scalar c. Now, for any phases $e^{i\theta}$ and $e^{i\phi}$, we have $f(e^{i\theta})f(e^{i\phi}) = f(e^{i(\theta+\phi)})$. Since $\hat{f}(1) = 1$ this implies $g(\theta)g(\phi) = g(\theta+\phi)$, i.e. g must be a real one-dimensional representation of $U(1)$. The only such representation is $g = 1$, hence $f(\alpha) = f(|\alpha|)$.

Now we will show that $f(x) = x^p$ for some p. Consider any $0 < x < 1$ and $0 < x' < 1$ where $x \neq x'$. Since $f(\alpha)f(\beta) = f(\alpha\beta)$, we must have that $f(x^k) = f(x)^k$ and $f(x'^k) = f(x')^k$ for any $k \in \mathbb{N}$. Let $p = \log(f(x))/\log(x)$ and $p' = \log(f(x'))/\log(x')$. Then the above equations imply that $f(x^k) = x^{kp}$ and $f(x'^k) = x'^{kp'}$ for all $k \in \mathbb{N}$.

Now suppose by way of contradiction that there exist x, x' such that $p \neq p'$. Since both $x < 1$ and $x' < 1$, as $k \to \infty$ we have that $f(x^{kp}) \to 0$ and $f(x'^{kp}) \to 0$. However, the sequence of points $f(x), f(x^2), f(x^3), \ldots$ approaches zero along the curve $h(x) = x^p$ while the sequence of points $f(x'), f(x'^2), f(x'^3), \ldots$ approaches zero along the curve $h'(x) = x^{p'}$. This implies f must oscillate infinitely many times between the curves h and h', which implies f' must change signs infinitely many times by the intermediate value theorem. This contradicts the Well-behaved derivative assumption.

Hence we have for all $0 < x < 1$, $f(x) = x^p$ for some p. Now if $x > 1$, we have $f(x)f(1/x) = f(1) = 1$. Since $1/x < 1$, then we have $f(1/x) = 1/x^p$, so $f(x) = x^p$ as well. Also $f(1) = 1^p = 1$, and by continuity $f(0) = 0$. Hence for all $x \geq 0$ we must have $f(x) = x^p$ for some p, as claimed. □

9.5.2 Born rule violations imply signaling and super-Grover speedup

We first show that large violations of the Born rule imply a large speedup to Grover search and allow for large amounts of superluminal signaling. This was previously shown by Aaronson [6], but for completeness we will summarize the proof here.

Theorem 9.5.2 (Aaronson [6] Theorem 6). *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$ where $\delta \neq 0$. Then one can solve PP problems on instances of size n in time $O(\frac{n^2}{|\delta|})$. In particular one can search an unordered list of 2^n indices in $O(\frac{n^2}{|\delta|})$ time.*

Proof. We will use the modified Born rule to simulate postselection. Suppose one has a state $|\Psi\rangle = \sum_x (\alpha_x|0\rangle + \beta_x|1\rangle)|x\rangle$ and wishes to simulate postselection of the first qubit on the state $|0\rangle$. Suppose $\delta > 0$; the case $\delta < 0$ follows analogously. To simulate postselection on zero, simply append k ancilla qubits in the $|0\rangle$ state. Then apply a Hadamard to each of

the ancilla qubits controlled on the first qubit being a 1. The state now evolves to

$$\sum_x \left(\alpha_x |0\rangle|x\rangle|0\rangle^{n/\delta} + \beta_x |1\rangle|x\rangle \sum_y 2^{-k/2}|y\rangle \right) \tag{9.117}$$

When measuring this state in the computational basis, the probability of measuring a 0 on the first qubit is proportional to $\sum_x |\alpha_x|^{2+\delta}$, while the probability of getting a 1 on the first qubit is proportional to $2^{-k\delta/2} \sum_x |\beta_x|^{2+\delta}$. Hence setting $k = n/\delta$, the probability of getting a 1 on the first qubit is exponentially suppressed by a factor of 2^{-n}. This effectively postselects the first qubit to have value 0 as desired. The rest of the proof follows from the fact that Aaronson's PostBQP algorithm to solve PP-hard problems on instances of size n runs in time $O(n)$ and involves $O(n)$ postselections; hence using this algorithm to solve PP-hard problems when the Born rule is violated takes time $O(\frac{n^2}{\delta})$ as claimed. □

Aaronson's result also implies that large violations of the Born rule imply one can send superluminal signals with small numbers of qubits.

Theorem 9.5.3 (Aaronson [6]). *Suppose that the Born rule is modified such that* $f(\alpha) = |\alpha|^{2+\delta}$ *where* $\delta \neq 0$. *Then one can transmit a bit superluminally in a protocol involving a state on* $O(n/|\delta|)$ *qubits which succeeds with probability* $1 - 2^{-n}$. *Note one can use this protocol to send either classical bits or quantum bits.*

Proof. The proof follows almost immediately from the proof of Theorem 9.5.2. Suppose that Alice wishes to send a bit 0 or 1 to Bob. Alice and Bob can perform the standard teleportation protocol [57], but instead of Alice sending her classical measurement outcomes to Bob, Alice simply postselects her measurement outcome to be 00 (i.e. no corrections are necessary to Bob's state) using the trick in Theorem 9.5.2. If Alice uses $O(n/|\delta|)$ qubits to simulate the postselection, and then measures her qubits, she will obtain outcome 00 with probability $1 - 2^{-n}$ and the bit will be correctly transmitted as desired. □

(Continues on next page.)

9.5.3 Signaling implies large power law violation

We now show that if one can send a superluminal signal with bias ε using a shared state on n qubits, then the violation of the Born rule δ must satisfy $|\delta| \geq O(\varepsilon/n)$. Hence δ and ε must be polynomially related. Put less precisely, if a physically reasonable experiment can send a superluminal signal with a nontrivial probability, then there must be a nontrivial (and hence observable) violation of the Born rule. This in turn, implies by Theorem 9.5.2 that one can solve NP-hard problems with a reasonable multiplicative overhead.

Theorem 9.5.4. *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$, and suppose there is a signaling protocol using an entangled state on n qubits signaling with probability ε. Then $|\delta| \geq O(\frac{\varepsilon}{n})$.*

Proof. Consider the most general signaling protocol to send a bit of information. Suppose that Alice and Bob share an entangled state $|\Phi\rangle$ on n qubits, m of which are held by Bob and $n - m$ of which are held by Alice. To send a zero, Alice performs some unitary U_0 on her half of the state, and to send a one, Alice performs some unitary U_1 on her half of the state. Bob then measures in some fixed basis B. This is equivalent to the following protocol: Alice and Bob share the state $|\Psi\rangle = U_0|\Phi\rangle$ ahead of time, and Alice does nothing to send a 0, and applies $U = U_1 U_0^\dagger$ to obtain $|\Psi'\rangle = U|\Psi\rangle$ send a 1. Then Bob measures in basis B. We say the protocol succeeds with probability ε if the distributions seen by Bob in the case Alice is sending a 0 vs. a 1 differ by ε in total variation distance. As shown in section 9.4.5, the total variation distance is polynomially related to the capacity of the resulting classical communication channel.

Let α_{xy} be the amplitude of the state $|x\rangle|y\rangle$ in $|\Psi\rangle$, where the $|x\rangle$ is an arbitrary basis for Alice's qubits and $|y\rangle$ are given by the basis B in which Bob measures his qubits. Let α'_{xy} be the amplitude of $|x\rangle|y\rangle$ in the state $|\Psi'\rangle$, so we have $\alpha'_{xy} = \sum_{x'} U_{xx'}\alpha_{x'y}$. In short

$$|\Psi\rangle = \sum_{xy} \alpha_{xy}|x\rangle|y\rangle \qquad |\Psi'\rangle = \sum_{xy} \alpha'_{xy}|x\rangle|y\rangle \tag{9.118}$$

Assume that $\sum_{x,y} |\alpha_{xy}|^2 = 1$, i.e. the state is normalized in the ℓ_2 norm. Since U is unitary this implies the state $U|\psi'\rangle$ is normalized in the ℓ_2 norm as well.

Now suppose that the protocol has an ε probability of success. Let D_0 be the distribution on outcomes $y \in \{0,1\}^m$ when Alice is sending a zero, and D_1 be the distribution when Alice is sending a 1. Let $D_b(y)$ denote the probability of obtaining outcome y under D_b. Then the total variation distance between D_0 and D_1, given by $\frac{1}{2}\sum_y |D_0(y) - D_1(y)|$, must be at least ε. Equivalently, there must be some event $S \subset \{0,1\}^m$ for which

$$\sum_{y \in S} D_0(y) - D_1(y) \geq \varepsilon \tag{9.119}$$

and for which, for all $y \in S$, we have $D_0(y) > D_1(y)$.

Assume for the moment that $\delta > 0$; an analogous proof will hold in the case $\delta < 0$. Let $N = 2^n$ be the dimension of the Hilbert space of $|\Psi\rangle$. Plugging in the probabilities $D_0(y)$

and $D_1(y)$ given by the modified Born rule, we obtain

$$\varepsilon \leq \sum_{x\in\{0,1\}^{n-m}, y\in S} \frac{|\alpha_{xy}|^{2+\delta}}{\sum_{x'y'}|\alpha_{x'y'}|^{2+\delta}} - \frac{|\alpha'_{xy}|^{2+\delta}}{\sum_{x'y'}|\alpha'_{x'y'}|^{2+\delta}} \tag{9.120}$$

$$\leq \sum_{x\in\{0,1\}^{n-m}, y\in S} N^{\delta/2}|\alpha_{xy}|^{2+\delta} - |\alpha'_{xy}|^{2+\delta} \tag{9.121}$$

$$= \sum_{x\in\{0,1\}^{n-m}, y\in S} \left(1+\frac{\delta}{2}\log(N)\right)|\alpha_{xy}|^2\left(1+\delta\log|\alpha_{xy}|\right) - |\alpha'_{xy}|^2(1+\delta\log|\alpha'_{xy}|) + O(\delta^2) \tag{9.122}$$

$$= \sum_{x\in\{0,1\}^{n-m}, y\in S} \left(|\alpha_{xy}|^2 - |\alpha'_{xy}|^2\right) + \frac{\delta}{2}\log(N)|\alpha_{xy}|^2 + \delta\left(|\alpha_{xy}|^2\log|\alpha_{xy}| - |\alpha'_{xy}|^2\log|\alpha'_{xy}|\right) \tag{9.123}$$

$$+ O(\delta^2)$$

$$\leq \frac{\delta}{2}\log(N) + \frac{\delta}{2}\sum_{x\in\{0,1\}^{n-m}, y\in S}\left(|\alpha_{xy}|^2\log|\alpha_{xy}|^2 - |\alpha'_{xy}|^2\log|\alpha'_{xy}|^2\right) + O(\delta^2) \tag{9.124}$$

$$\leq \frac{\delta}{2}\log(N) + \frac{\delta}{2}\log(N) + O(\delta^2) = \delta n + O(\delta^2) \tag{9.125}$$

On line (9.121) we used the fact that for any vector $|\phi\rangle = \sum_y \beta_y|y\rangle$ of ℓ_2 norm 1 over a Hilbert space of dimension N, we have $N^{-\delta/2} \leq \sum_y |\beta_y|^{2+\delta} \leq 1$ when $\delta > 0$. On line (9.122) we expanded to first order in δ. On line (9.124) we used the fact that the first term is zero because applying a unitary to one half of a system does not affect measurement outcomes on the other half of the system and the second sum is upper bounded by 1. On line (9.125) we used the fact that the sum is given by a difference of entropies of (possibly subnormalized) probability distributions, each of which is between zero and $\log(N)$.

Hence we have that $\delta n + O(\delta^2) \geq \varepsilon$, so to first order in δ we must have $\delta \geq \varepsilon/n$ as claimed. □

The following corollary follows from Theorem 9.5.2, and hence we've shown that superluminal signaling implies a super-Grover speedup.

Corollary 9.5.1. *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$, and that there is a signaling protocol using an entangled state on n qubits which signals with probability ε. Then there is an algorithm to solve #P-hard and NP-hard instances of size m (e.g. #SAT on m variables) in time $O(m^2 n/\varepsilon)$.*

9.5.4 Super-Grover speedup implies signaling

We now show that even a mild super-Grover speedup implies that δ is large, and hence one can send superluminal signals. Our proof uses the hybrid argument of Bennett, Bernstein, Brassard and Vazirani [56] combined with the proof techniques of Theorem 9.5.4.

Theorem 9.5.5. *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$, and there is an algorithm to search an unordered list of N items with Q queries using an algorithm*

over a Hilbert space of dimension M. Then

$$\frac{1}{6} \leq \frac{2Q}{\sqrt{N}} + |\delta| \log(M) + O(\delta^2). \tag{9.126}$$

Proof. Suppose that such an algorithm exists. It must consist of a series of unitaries and oracle calls followed by a measurement in the computational basis.

Let $|\psi^0\rangle = \sum_y \alpha_y^0 |y\rangle$ be the state of the algorithm just before the final measurement when there is no marked item, and let $|\psi^x\rangle = \sum_y \alpha_y^x |y\rangle$ be the state if there is a marked item. Let D_0 be the distribution on y obtained by measuring $|\psi^0\rangle$ in the computational basis, and D_x be the distribution obtained by measuring $|\psi^x\rangle$. We know that $|\psi^0\rangle$ and $|\psi^x\rangle$ must be distinguishable with $2/3$ probability for every x. Hence we must have that the total variation distance between D_0 and D_x must be at least $1/6$ for every x (otherwise one could not decide the problem with bias $1/6$). This implies there must exist some event S_x for which

$$\frac{1}{6} \leq \sum_{y \in S_x} D_0(y) - D_1(y) \tag{9.127}$$

Assume $\delta > 0$; an analogous proof holds for $\delta < 0$. Plugging in the expressions for D_0 and D_1 and averaging over x we obtain

$$\frac{1}{6} \leq \frac{1}{N} \sum_x \sum_{y \in S_x} \frac{|\alpha_y^0|^{2+\delta}}{\sum_y |\alpha_{y'}^0|^{2+\delta}} - \frac{|\alpha_y^x|^{2+\delta}}{\sum_y |\alpha_{y'}^x|^{2+\delta}} \tag{9.128}$$

$$\leq \frac{1}{N} \sum_x \sum_{y \in S_x} M^{\delta/2} |\alpha_y^0|^{2+\delta} - |\alpha_y^x|^{2+\delta} \tag{9.129}$$

$$= \frac{1}{N} \sum_x \sum_{y \in S_x} \left(1 + \frac{\delta}{2} \log(M)\right) |\alpha_y^0|^2 \left(1 + \delta \log |\alpha_y^0|\right) - |\alpha_y^x|^2 \left(1 + \delta \log |\alpha_y^x|\right) + O\left(\delta^2\right) \tag{9.130}$$

$$= \frac{1}{N} \sum_x \sum_{y \in S_x} \left(|\alpha_y^0|^2 - |\alpha_y^x|^2\right) + \frac{\delta}{2} \log(M) |\alpha_y^0|^2 + \frac{\delta}{2} \left(|\alpha_y^0|^2 \log |\alpha_y^0|^2 - |\alpha_y^x|^2 \log |\alpha_y^x|^2\right) \tag{9.131}$$

$$+ O\left(\delta^2\right)$$

$$\leq \delta \log(M) + O(\delta^2) + \frac{1}{N} \sum_x \sum_{y \in S_x} \left(|\alpha_y^0|^2 - |\alpha_y^x|^2\right). \tag{9.132}$$

In line (9.129) we used the fact that $M^{-\delta/2} \leq \sum_y |\alpha_y|^{2+\delta} \leq 1$ for any state α_y normalized in the ℓ_2 norm, in line (9.130) we Taylor expanded to first order in δ, and in line (9.132) we used the fact that the sum in the second term is upper bounded by one and the sum on third term is a difference of entropies of (subnormalized) probability distributions which is at most $\log(M)$.

We next consider the final term

$$R \equiv \frac{1}{N} \sum_x \sum_{y \in S_x} \left(|\alpha_y^0|^2 - |\alpha_y^x|^2 \right). \tag{9.133}$$

Let $\hat{S}_x$ be the observable

$$\hat{S}_x = \sum_{y \in S_x} |y\rangle\langle y|. \tag{9.134}$$

Then

$$\begin{aligned} R &= \frac{1}{N} \sum_x \left[\langle \psi^0 | \hat{S}_x | \psi^0 \rangle - \langle \psi^x | \hat{S}_x | \psi^x \rangle \right] && (9.135) \\ &= \frac{1}{N} \sum_x \left[\left(\langle \psi^0 | - \langle \psi^x | \right) \hat{S}_x | \psi^0 \rangle + \langle \psi^x | \hat{S}_x \left(|\psi^0\rangle - |\psi^x\rangle \right) \right] && (9.136) \\ &\leq \frac{2}{N} \sum_x \left\| |\psi^0\rangle - |\psi^x\rangle \right\|, && (9.137) \end{aligned}$$

where the last inequality uses the fact that $\|\hat{S}_x\| = 1$. Next we note that $\sum_x \left\| |\psi^0\rangle - |\psi^x\rangle \right\|$ is the ℓ_1 norm of the N-dimensional vector whose x^{th} component is $\left\| |\psi^0\rangle - |\psi^x\rangle \right\|$. For any N-dimensional vector $\vec{v}$, $\|v\|_1 \leq \sqrt{N} \|\vec{v}\|_2$. Thus,

$$R \leq \frac{2}{\sqrt{N}} \sqrt{\sum_x \| |\psi^0\rangle - |\psi^x\rangle \|^2}. \tag{9.138}$$

As shown in [56], a unitary search algorithm using Q oracle queries yields

$$\sum_x \left\| |\psi^0\rangle - |\psi^x\rangle \right\|^2 \leq 4Q^2. \tag{9.139}$$

Together, Eq. (9.138) and Eq. (9.139) imply

$$R \leq \frac{2Q}{\sqrt{N}}. \tag{9.140}$$

Now, Eq. (9.140) bounds the last term in Eq. (9.132) yielding our final result.

$$\frac{1}{6} \leq \delta \log(M) + O(\delta^2) + \frac{2Q}{\sqrt{N}}. \tag{9.141}$$

□

The following Corollary follows immediately from Theorem 9.5.5 and Theorem 9.5.3.

Corollary 9.5.2. *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$, and one can search a list of $N = 2^n$ items using m qubits and Q queries. Then to first order in δ, we have*

$$|\delta| \geq \frac{1}{m} \left(\frac{1}{6} - \frac{2Q}{\sqrt{N}} \right).$$

In particular, if one can search an N element list with $Q \leq \sqrt{N}/24$ queries on a state of m

qubits, then $|\delta| \geq \frac{1}{12m}$, *and hence by Theorem 9.5.3 one can send superluminal signals with probability 2/3 using* $O(m)$ *qubits.*

In contrast, Grover's algorithm uses $\frac{\pi}{4}\sqrt{N}$ queries to solve search, which is optimal [246]. So Corollary 9.5.2 shows that if one can achieve even a modest factor of ($6\pi \approx 19$) speedup over Grover search using m qubits, then one can send superluminal signals using $O(m)$ qubits.

9.6 Proofs: Cloning of Quantum States

One way to modify quantum mechanics is to allow perfect copying of quantum information, or "cloning". As a minimal example, we will here introduce the ability to do perfect single-qubit cloning. As with nonlinear dynamics, care must be taken to formulate a version of quantum cloning that is actually well defined. It is clear that perfect single qubit cloning should take $|\psi\rangle \mapsto |\psi\rangle \otimes |\psi\rangle$ for any single-qubit pure state. The nontrivial task is to define the behavior of the cloner on qubits that are entangled. It is tempting to simply define cloning in terms of the Schmidt decomposition of the entangled state. That is, applying the cloner to qubit B induces the map $\sum_i \lambda_i |i_A\rangle|i_B\rangle \mapsto \sum_i \lambda_i |i_A\rangle|i_B\rangle|i_B\rangle$. However, this prescription is ill-defined due to the non-uniqueness of Schmidt decompositions. The two decompositions of the EPR pair given in Eq. (9.163) and Eq. (9.164) provide an example of the inconsistency of the above definition.

Instead, we define our single-qubit cloner as follows.

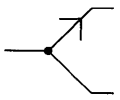
Definition 9.6.1. *Let* ρ_{AB} *be a state on a bipartite system* AB. *Let* ρ_B *be the reduced density matrix of* B. *Then applying the cloner to* B *yields*

$$\rho_{AB} \mapsto \rho_{AB} \otimes \rho_B.$$

In particular, for pure input, we have $|\psi_{AB}\rangle\langle\psi_{AB}| \mapsto |\psi_{AB}\rangle\langle\psi_{AB}| \otimes \rho_B$. Thus, this version of cloning maps pure states to mixed states in general. Furthermore, the clones are asymmetric. The cloner takes one qubit as input and produces two qubits as output. The two output qubits have identical reduced density matrices. However, one of the output qubits retains all the entanglement that the input qubit had with other systems, whereas the other qubit is unentangled with anything else. By monogamy of entanglement it is impossible for both outputs to retain the entanglement that the input qubit had.

It is worth noting that the addition of nonlinear dynamics, and cloning in particular, breaks the equivalence between density matrices and probabilistic ensembles of pure states. Here, we take density matrices as the fundamental objects in terms of which our generalized quantum mechanics is defined.

In analyzing a model of computation involving cloning, we will treat the cloning operation as an additional gate, with the same "cost" as any other. In circuit diagrams, we denote the cloning gate as follows.



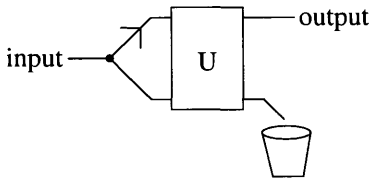
This notation reflects the asymmetric nature of our cloning gate; the arrow indicates the output qubit that retains the entanglement of the input qubit.

9.6.1 Grover Search using Quantum Cloning

Cloning is a nonlinear map on quantum states. As argued by Abrams and Lloyd [22], one can solve Grover search on a database of size N using $O(1)$ oracle queries and $O(\log N)$ applications of S, for any nonlinear map S from pure states to pure states, except perhaps some pathological cases. Here, with theorem 9.8.1, we have formalized this further, showing that this holds as long as S is differentiable. However, the cloning gate considered here maps pure states to mixed states. Therefore, this gate requires a separate analysis. We cannot simply invoke theorem 9.8.1. Instead we specifically analyze the cloning gate given above and arrive at the following result.

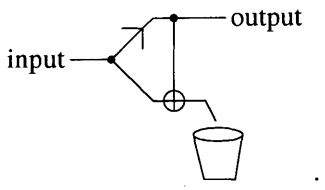
Theorem 9.6.1. *Suppose we have access to a standard Grover bit-flip oracle, which acts as* $U_f|y\rangle|z\rangle = |y\rangle|z \oplus f(y)\rangle$ *where* $f : \{0,1\}^n \to \{0,1\}$. *Using one query to this oracle, followed by a circuit using* poly(n) *conventional quantum gates and* $O(n)$ *of the single-qubit cloning gates described in definition 9.6.1, one can distinguish between the cases that* $|f^{-1}(1)| = 0$ *and* $|f^{-1}(1)| = 1$ *with high probability.*

Proof. For the design of nonlinear Grover search algorithms it is helpful to have a nonlinear map from a fixed state space to itself. To this end, we consider circuits of the following form, which implement nonlinear maps from the space of possible density matrices of a qubit to itself.



Here, one clones the input qubit, performs some unitary U between the two resulting copies, and lastly discards one of the qubits.

With a small amount of trial and error one can find a choice of U which enables single-query Grover search using an analogue of the Abrams-Lloyd algorithm. Specifically, we choose U to be the controlled-not gate. That is, let

$\mathcal{M} =$ 

$\mathcal{M}$ is a quadratic map on density matrices. By direct calculation

$$\mathcal{M}\left(\begin{bmatrix} r_{00} & r_{01} \\ r_{10} & r_{11} \end{bmatrix}\right) = \begin{bmatrix} r_{00}^2 + r_{00}r_{11} & r_{01}^2 + r_{01}r_{10} \\ r_{10}^2 + r_{10}r_{01} & r_{11}^2 + r_{11}r_{00} \end{bmatrix}. \tag{9.142}$$

One can find the fixed points of $\mathcal{M}$ by solving the system of four quadratic equations

implied by $\mathcal{M}(\rho) = \rho$. The solutions are as follows.

$$r_{10} = 1 - r_{01}, \quad r_{11} = 1 - r_{00} \tag{9.143}$$

$$r_{00} = 0, \quad r_{10} = 1 - r_{01}, \quad r_{11} = 0 \tag{9.144}$$

$$r_{01} = 1, \quad r_{10} = 0, \quad r_{11} = 1 - r_{00} \tag{9.145}$$

$$r_{00} = r_{01} = r_{10} = r_{11} = 0 \tag{9.146}$$

The solutions Eq. (9.144) and Eq. (9.146) are traceless and therefore unphysical. Solution Eq. (9.145) is an arbitrary mixture of $|0\rangle$ and $|1\rangle$. That is,

$$\rho_r = \begin{bmatrix} r & 0 \\ 0 & 1-r \end{bmatrix} \quad \text{is a fixed point for all } r \in [0,1]. \tag{9.147}$$

As a matrix, the solution Eq. (9.143) is

$$\rho_{a,b} = \begin{bmatrix} a & b \\ 1-b & 1-a \end{bmatrix}. \tag{9.148}$$

This is only Hermitian if $b = (1-b)^*$, which implies that $b = \frac{1}{2} + \alpha i$ for some $\alpha \in \mathbb{R}$. However, if $\alpha \neq 0$ then $\rho_{a,b}$ fails to be positive semidefinite, which is unphysical. Thus, $b = \frac{1}{2}$. The eigenvalues of $\rho_{a,1/2}$ are

$$\frac{1}{2} \pm \sqrt{\frac{1 + 2a(a-1)}{2}}. \tag{9.149}$$

Thus, unless $a = \frac{1}{2}$, the largest eigenvalue of $\rho_{a,1/2}$ exceeds one, which is unphysical. So, the only physical fixed point other than ρ_r is

$$\rho_+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = |+\rangle\langle+|. \tag{9.150}$$

Numerically, one finds that ρ_r is an attractive fixed point and ρ_+ is a repulsive fixed point. Let

$$\rho_\varepsilon = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} - \varepsilon \\ \frac{1}{2} - \varepsilon & \frac{1}{2} \end{bmatrix} = (1-\varepsilon)|+\rangle\langle+| + \varepsilon|-\rangle\langle-|. \tag{9.151}$$

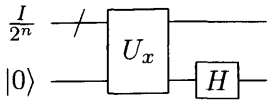
Then,

$$\mathcal{M}(\rho_\varepsilon) = \rho_{2\varepsilon + O(\varepsilon^2)}. \tag{9.152}$$

Consequently, $\mathcal{M}^r(\rho_\varepsilon)$ is easily distinguishable from $\mathcal{M}^r(\rho_+) = \rho_+$ after $r = O(\log(1/\varepsilon))$ iterations of $\mathcal{M}$.

Let U_f be the standard Grover bit-flip oracle, which acts as $U_f|y\rangle|z\rangle = |y\rangle|z \oplus f(y)\rangle$

where $f : \{0,1\}^n \to \{0,1\}$. Now, consider the following circuit.



One sees that the bottom qubit emerges in the state ρ_+ if f has no solution and emerges in the state ρ_ε with $\varepsilon = \frac{1}{2^n}$ if f has one solution. By making one such query and then applying the map $\mathcal{M}$ a total of $O(n)$ times to the resulting state, one obtains single-qubit states in the no-solution and one-solution cases that are easily distinguished with high confidence using conventional quantum measurements. □

For simplicity, in theorem 9.6.1, we have restricted our attention to search problems which are promised to have exactly one solution or no solutions and our task is to determine which of these is the case. Note that 3SAT can be reduced to UNIQUESAT in randomized polynomial time [235]. Hence solving the Grover problem in poly(n) time when there is either exactly one solution or no solutions suffices to solve NP-hard problems in randomized polynomial time.

It is interesting to note that probability distributions also cannot be cloned. The map $\vec{p} \mapsto \vec{p} \otimes \vec{p}$ on vectors of probabilities is nonlinear and hence does not correspond to any realizable stochastic process. Furthermore, one finds by a construction similar to the above that cloning of classical probability distributions also formally implies polynomial-time solution to NP-hard problems via logarithmic-complexity single-query Grover search. However, nonlinear maps on probabilities do not appear to be genuinely well-defined. Suppose we have probability p_1 of drawing from distribution $\vec{p_1}$ and probability p_2 of drawing from distribution $\vec{p_2}$. Normally this is equivalent to drawing from $p_1\vec{p_1} + p_2\vec{p_2}$. However, if we apply a nonlinear map $\mathcal{M}$ then $\mathcal{M}(p_1\vec{p_1} + p_2\vec{p_2})$ is in general not equal to $p_1\mathcal{M}(p_1) + p_2\mathcal{M}(\vec{p_2})$. It is not clear that a well-defined self-consistent principle can be devised for resolving such ambiguities.

9.6.2 Superluminal Signaling using Quantum Cloning

Suppose Alice and Bob share an EPR pair $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. If Alice wishes to transmit a zero she does nothing. If she wishes to transmit a one she measures her qubit in the computational basis. If Alice doesn't measure then Bob's reduced density matrix is maximally mixed. Hence if he makes several clones and measures them all in the computational basis he will obtain a uniformly random string of ones and zeros. If Alice does measure then Bob's reduced density matrix is either $|0\rangle\langle 0|$ or $|1\rangle\langle 1|$, with equal probability. If he makes several clones and measures them all in the computational basis he will get $000\ldots$ or $111\ldots$, with equal probability. Thus, by making logarithmically many clones, Bob can achieve polynomial certainty about the bit that Alice wished to transmit.

9.7 Proofs: Postselection

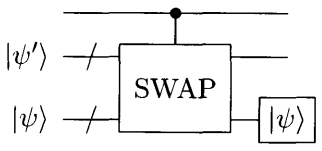
In [6] it was shown that adding the ability to postselect a single qubit onto the state $|0\rangle$ to the quantum circuit model yields a model of computation whose power is equal to the classical complexity class PP. Furthermore, postselection onto $|0\rangle$ allows perfect superluminal

signaling by postselected quantum teleportation. Here we consider a more general question: suppose we have the ability to postselect on some arbitrary but fixed n-qubit state $|\psi\rangle$. Does this still yield efficient means of solving problems in PP and sending superluminal signals? It is clear that one can use postselection onto $|\psi\rangle$ to simulate postselection onto $|0\rangle$ given a quantum circuit for a unitary U such that $U|00\ldots0\rangle = |\psi\rangle$. However, for a generic n-qubit state $|\psi\rangle$, no polynomial-size quantum circuit for this tasks exists. Nevertheless, in this section we show that, for Haar random (but fixed) $|\psi\rangle$, postselection onto $|\psi\rangle$ can with high probability be used to simulate postselection onto $|0\rangle$ with exponential precision.

We first note that the maximally entangled state of $2n$ qubits:

$$|\Phi_{2n}\rangle = \sum_{x\in\{0,1\}^n} |x\rangle \otimes |x\rangle \tag{9.153}$$

can be prepared using n Hadamard gates followed by n CNOT gates. Postselecting the second tensor factor of $|\Phi_{2n}\rangle$ onto $|\psi\rangle$ yields $|\psi\rangle$ on the first tensor factor. In this manner, one may extract a copy of $|\psi\rangle$. We assume that $|\psi\rangle$ is Haar random but fixed. That is, each time one uses the postselection "gate," one postselects onto the same state $|\psi\rangle$. Hence, using the above procedure twice yields two copies of $|\psi\rangle$. Applying σ_x to one qubit of one of the copies of $|\psi\rangle$ yields a state $|\psi'\rangle = \sigma_x|\psi\rangle$. As shown below, the root-mean-square inner product between $|\psi\rangle$ and $|\psi'\rangle$ is of order $1/\sqrt{2^n}$. That is, they are nearly orthogonal. Thus, one can simulate postselection onto $|0\rangle$ with the following circuit.



Here, the top qubit gets postselected onto $|0\rangle$ with fidelity $1 - O(1/\sqrt{2^n})$, the middle register is discarded, and the bottom register is postselected onto $|\psi\rangle$, an operation we denote by —|ψ⟩—.

Lastly, we prove the claim that the root-mean-square inner product between a Haar random n-qubit state $|\psi\rangle$ and $|\psi'\rangle = \sigma_x|\psi\rangle$ is of order $1/\sqrt{2^n}$. This mean-square inner product can be written as

$$\bar{I} = \int dU |\langle 0\ldots0|U^\dagger \sigma_x U|0\ldots0\rangle|^2 \tag{9.154}$$

$$= \sum_{a,b\in\{0,1\}^n} \int dU U^\dagger_{0a} U_{\bar{a}0} U^\dagger_{0b} U_{\bar{b}0} \tag{9.155}$$

where $\bar{a}$ indicates the result of flipping the first bit of a, $\bar{b}$ indicates the result of flipping the first bit of b, and 0 in the subscripts is shorthand for the bit string $0\ldots0$. (We arbitrarily choose the σ_x to act on the first qubit.)

Next we recall the following identity regarding integrals on the Haar measure over $U(N)$. (See [93] or appendix D of [148].)

$$\begin{aligned}\int dU\ U_{ij}U_{kl}U^\dagger_{mn}U^\dagger_{op} = &\frac{1}{N^2-1}\left(\delta_{in}\delta_{kp}\delta_{jm}\delta_{lo} + \delta_{ip}\delta_{kn}\delta_{jo}\delta_{lm}\right)\\ &-\frac{1}{N(N^2-1)}\left(\delta_{ij}\delta_{kp}\delta_{jo}\delta_{lm} + \delta_{ip}\delta_{kn}\delta_{jm}\delta_{lo}\right)\end{aligned} \tag{9.156}$$

Applying Eq. (9.156) to Eq. (9.155) shows that the only nonzero terms come from $a = \bar{b}$ and consequently

$$\bar{I} = \sum_{a \in \{0,1\}^n} \int dU \; U_{\bar{a}0} U_{a0} U_{0a}^\dagger U_{0\bar{a}}^\dagger \tag{9.157}$$

$$= \frac{N}{N^2 - 1} - \frac{1}{N^2 - 1}. \tag{9.158}$$

Consequently, the RMS inner product for large N is

$$\sqrt{\bar{I}} \simeq \frac{1}{\sqrt{N}}. \tag{9.159}$$

Recalling that $N = 2^n$ completes the argument.

9.8 Proofs: General Nonlinearities

Our discussion of final-state projection models can be thought of as falling within a larger tradition of studying the information-theoretic and computational complexity implications of nonlinear quantum mechanics, as exemplified by [22, 4, 189, 91]. A question within this subject that has been raised multiple times [22, 4] is whether all nonlinearities necessarily imply that Grover search can be solved with a single query. In this note we shed some light on this question. However, note that the setting differs from that of section 9.4.3 in that (following [22, 4]) we assume the nonlinear map is the same each time, and we can apply it polynomially many times. In section 9.4.3 we have included the possibility that black holes (and the nonlinear maps that they generate) are scarce and that they may differ from one another.

We first note that, for dynamics that map normalized pure states to normalized pure states, the terms nonunitary and nonlinear are essentially interchangeable. Let V be the manifold of normalized vectors on a complex Hilbert space $\mathcal{H}$, which could be finite dimensional or infinite-dimensional. Let $S : V \to V$ be a general map, not necessarily linear or even continuous. We'll call S a *unitary map* if it preserves the magnitude of inner products. That is, $|\langle S\psi|S\phi\rangle| = |\langle\psi|\phi\rangle|$ for all $|\phi\rangle, |\psi\rangle \in \mathcal{H}$. Wigner's theorem [242] states that all unitary maps are either unitary linear transformations, or antiunitary antilinear transformations. (Antiunitary transformations are equivalent to unitary transformations followed by complex conjugation of all amplitudes in some basis.) Extending quantum mechanics by allowing antiunitary dynamics does not affect computational complexity, as can be deduced from [27]. Thus, without loss of generality, we may ignore antiunitary maps. Hence, within the present context, if a map is unitary then it is linear. Conversely, by linear algebra, if map S is linear, and maps $V \to V$, *i.e.* is norm-preserving, then it is also inner-product preserving, *i.e.* unitary.

A standard version of the Grover problem is, for some function $f : \{0,1\}^n \to \{0,1\}$, to decide whether the number of solutions to $f(y) = 1$ is zero or one, given that one of these is the case. The search problem of finding a solution is reducible to this decision problem with logarithmic overhead via binary search. In [22] Abrams and Lloyd show how to solve the decision version of Grover search using a single quantum query to f and $O(n)$ applications of a single-qubit nonlinear map. This suffices to solve NP in polynomial time. We now briefly describe their algorithm. In contrast to section 9.4.3, it is more convenient here to assume

a bit-flip oracle rather than a phase-flip oracle. That is, for $y \in \{0,1\}^n$ and $z \in \{0,1\}$ the oracle O_f acts as

$$O_f|y\rangle|z\rangle = |y\rangle|z \oplus f(y)\rangle. \tag{9.160}$$

Querying the oracle with the state $\frac{1}{\sqrt{2^n}} \sum_{y\in\{0,1\}^n} |y\rangle|0\rangle$ yields $\frac{1}{\sqrt{2^n}} \sum_{y\in\{0,1\}^n} |y\rangle|f(y)\rangle$. Applying a Hadamard gate to each qubit of the first register and measuring the first register in the computational basis yields the outcome $00\ldots0$ with probability at least $\frac{1}{4}$. Given that this occurs, the post-measurement state of the second register is

$$|\psi_s\rangle = \frac{(2^n - s)|0\rangle + s|1\rangle}{\sqrt{(2^n - s)^2 + s^2}}, \tag{9.161}$$

where s is the number of solutions, *i.e.* $s = |f^{-1}(1)|$. Thus, we can solve the Grover search problem by distinguishing two exponentially-close states, namely $|\psi_0\rangle$ and $|\psi_1\rangle$. For the particular nonlinear map on the manifold of normalized pure single-qubit states considered in [22], a pair of states ε-close together can be separated to constant distance by iterating the map $O(\log(1/\varepsilon))$ times.

We now show that any differentiable nonlinear map from pure states to pure states on any finite-dimensional Hilbert space can achieve this. (See theorem 9.8.1.) Let $V^{(n)}$ be the manifold of normalized pure states on $\mathbb{C}^n$. Thus, $V^{(n)}$ is a $2n-1$ dimensional real closed compact manifold. For points a, b on $V^{(n)}$ let $|a-b|$ denote their distance. (Our choice of distance metric is not important to the argument, but for concreteness, we could choose the angle between quantum states, that is, $|a-b| = \cos^{-1}|\langle a|b\rangle|$. That this is a metric is proven in section 9.2.2 of [199].)

Theorem 9.8.1. *Let $S : V^{(n)} \to V^{(n)}$ be a differentiable map, that is, a self-diffeomorphism of $V^{(n)}$. Let $r = \max_{a,b\in V^{(n)}} \frac{|S(a) - S(b)|}{|a-b|}$. Then there exists some sufficiently short geodesic l in $V^{(n)}$ such that for all $x, y \in l$, $\frac{|S(x) - S(y)|}{|x-y|} \geq r$.*

Proof. Choose two points x, y on $V^{(n)}$ that maximize the ratio $r = \frac{|S(x) - S(y)|}{|x-y|}$. By assumption, S is not unitary, so not all distances are preserved. Because S is a map from $V^{(n)}$ to another manifold of equal volume (namely $V^{(n)}$ itself) it cannot be that all distances are decreased. Thus, this maximum ratio must be larger than one. The extent that this ratio exceeds one quantifies the deviation from unitarity.

Now, consider the geodesic g on $V^{(n)}$ from x to y. Because it is a geodesic, g has length $|x-y|$. Now consider the image of g under the map S. Because S is a continuous map, $S(g)$ will also be a line segment. By the construction, the endpoints of $S(g)$ are distance $r|x-y|$ apart. Therefore, the length of $S(g)$, which we denote $|S(g)|$, satisfies $|S(g)| \geq r|x-y|$, with equality if $S(g)$ happens to also be a geodesic. Thus, S induces a diffeomorphism S_g from the line segment g to the line segment $S(g)$, where $|S(g)|/|g| \geq r$. Because S_g is a diffeomorphism it follows that on any sufficiently small subsegment of g it acts by linearly magnifying or shrinking the subsegment and translating to some location on $S(g)$. Because $|S(g)|/|g| \geq r$ it follows that there exists some subsegment l such that this linear magnification is by a factor of at least r. (There could be some subsegments that grow less than this or even shrink, but if so, others have to make up for it by growing by a factor of

more $|S(g)|/|g|$.) □

We now argue that the existence of l suffices to ensure success for the Abrams-Lloyd algorithm. Let f denote the "magnification factor" that S induces on l. According to theorem 9.8.1, $f \geq r$. We are interested in asymptotic complexity, so the distance ε between $|\psi_0\rangle$ and $|\psi_1\rangle$ is asymptotically small. Therefore, we assume ε is smaller than the length of l. So, we can append ancilla qubits and apply a unitary transformation such that the resulting isometry maps $|\psi_0\rangle$ and $|\psi_1\rangle$ to two points $\left|\phi_0^{(0)}\right\rangle$ and $\left|\phi_1^{(0)}\right\rangle$ that lie on l. We then apply S, resulting in the states $\left|\phi_0^{(1)}\right\rangle$ or $\left|\phi_1^{(1)}\right\rangle$, which have distance $f\varepsilon$. If $f\varepsilon$ is larger than the length l then we terminate. Because we have a fixed nonunitary map, the distance between our states is now a constant (independent of ε and hence of the size of the search space). If $f\varepsilon$ is smaller than the length of l, then we apply a unitary map that takes $\left|\phi_0^{(1)}\right\rangle$ and $\left|\phi_1^{(1)}\right\rangle$ back onto l and apply S again. We then have states $\left|\phi_0^{(2)}\right\rangle$ and $\left|\phi_1^{(2)}\right\rangle$ separated by distance $f^2\varepsilon$. We then iterate this process until we exceed the size l, which separates the states to a constant distance and uses $\log_f(1/\varepsilon)$ of the nonunitary operations. States with constant separation can be distinguished within standard quantum mechanics by preparing a constant number of copies and collecting statistics on the outcomes of ordinary projective measurements.

9.9 A Cautionary Note on Nonlinear Quantum Mechanics

The Horowitz-Maldecena final-state projection model, cloning of quantum states, and the Gross-Pitaevsky equation (if interpreted as a quantum wave equation) all involve nonlinear dynamics of the wavefunction. In such cases, one must be very careful to ensure that subsystem structure, which is captured by tensor product structure in conventional quantum mechanics, is well-defined. Indeed, subsystem structure is lost by introducing generic nonlinearities, and in particular by the nonlinearity of the Gross-Pitaevsky equation. This makes the question about superluminal signaling in the Gross-Pitaevsky model ill-posed. The Horowitz-Maldecena model does have a natural notion of subsystem structure, which is one of the features that makes it appealing. Furthermore, the model of cloning that we formulate in Section 9.7 preserves subsystem structure by virtue of being phrased in terms of reduced density matrices.

More formally, let V be the manifold of normalized vectors in the Hilbert space $\mathbb{C}^d$. We will model nonlinear quantum dynamics by some map $S : V \to V$ which may not be a linear map on $\mathbb{C}^d$. In general, specifying a map S on V does not uniquely determine the action of S when applied to a subsystem of a larger Hilbert space. For example, consider the map S_0 on the normalized pure states of one qubit given by

$$S_0|\psi\rangle = |0\rangle \quad \forall|\psi\rangle \tag{9.162}$$

Now, consider what happens if we apply S_0 to half of an EPR pair $|\Psi_{\text{EPR}}\rangle$. We can write the EPR state in two equivalent ways

$$|\Psi_{\text{EPR}}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle|0\rangle + |1\rangle|1\rangle\right) \tag{9.163}$$

$$= \frac{1}{\sqrt{2}}\left(|+\rangle|+\rangle + |-\rangle|-\rangle\right) \tag{9.164}$$

where

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle) \tag{9.165}$$

Symbolically applying the rule $S_0|\psi\rangle \mapsto |0\rangle$ to the first tensor factor of Eq. (9.163) yields $|0\rangle|+\rangle$, whereas applying this rule to the first tensor factor of *Eq.*(9.164) yields $|0\rangle|0\rangle$.

This example illustrates that one must specify additional information beyond the action of a nonlinear map on a fixed Hilbert space in order to obtain a well-defined extension to quantum theory incorporating the notion of subsystems.

9.10 Open Problems

We have shown that in several domains of modifications of quantum mechanics, the resources required to observe superluminal signaling or a speedup over Grover's algorithm are polynomially related. We extrapolate that this relationship holds more generally, that is, in any quantum-like theory, the Grover lower bound is derivable from the no-signaling principle and vice-versa. A further hint in this direction is that, as shown in [44], the limit on distinguishing non-orthogonal states in quantum mechanics is dictated by the no-signaling principle. Thus, any improvement over the Grover lower bound based on beyond-quantum state discrimination can be expected to imply some nonzero capacity for superluminal signaling. There is a substantial literature on generalizations of quantum mechanics which could be drawn upon to address this question. In particular, one could consider the generalized probabilistic theories framework of Barrett [50], the category-theoretic framework of Abramsky and Coecke [23], the Newton-Schr'odinger equation [210], quaternionic quantum mechanics [25], or the Papadodimas-Raju state-dependence model of black hole dynamics [203, 186, 147]. In these cases the investigation of computational and communication properties is inseparably tied with the fundamental questions about the physical interpretations of these models. Possibly, such investigation could help shed light on these fundamental questions.

Our finding can be regarded as evidence against the possibility of using black hole dynamics to efficiently solve NP-complete problems, at least for problem instances of reasonable size. Note however that there are other independent questions regarding the feasibility of computational advantage through final-state projection and other forms of non-unitary quantum mechanics. In particular, the issue of fault-tolerance in modified quantum mechanics remains largely open, although some discussion of this issue appears in [78, 22, 4]. Also, while our results focus on the query complexity of search, in practice one also is interested in the time complexity. Harlow and Hayden [150] have argued that decoding the Hawking radiation emitted by a black hole may require exponential time on a quantum computer. If the Harlow-Hayden argument is correct, then exponential improvement in query complexity for search does not imply exponential improvement in time-complexity. We emphasize however that query complexity sets a lower bound on time complexity, and therefore the reverse implication still holds, namely exponential improvement in time complexity implies exponential improvement in query complexity, which in the models we considered implies superluminal signaling. Hence an operational version of the Grover lower bound can be derived from an operation version of the no-signaling principle.

Bibliography

[1] Scott Aaronson. Personal communication.

[2] Scott Aaronson. Quantum lower bound for the collision problem. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 635–642. ACM, 2002.

[3] Scott Aaronson. Is quantum mechanics an island in theoryspace? In *Proceedings of the Växjö Conference "Quantum Theory: Reconsideration of Foundations"*, 2004.

[4] Scott Aaronson. NP-complete problems and physical reality. *ACM SIGACT News*, 36(1):30–52, 2005. arXiv:quant-ph/0502072.

[5] Scott Aaronson. Quantum computing and hidden variables. *Physical Review A*, 71(3):032325, 2005.

[6] Scott Aaronson. Quantum computing, postselection, and probabilistic polynomial-time. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 461, pages 3473–3482. The Royal Society, 2005.

[7] Scott Aaronson. BQP and the polynomial hierarchy. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 141–150. ACM, 2010.

[8] Scott Aaronson. Impossibility of succinct quantum proofs for collision-freeness. *Quantum Information & Computation*, 12(1-2):21–28, 2012.

[9] Scott Aaronson. Postbqp postscripts: A confession of mathematical errors. 2014.

[10] Scott Aaronson. *P−?NP*, pages 1–122. Springer International Publishing, Cham, 2016.

[11] Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 333–342. ACM, 2011.

[12] Scott Aaronson, Adam Bouland, Lynn Chua, and George Lowther. ψ. *Phys. Rev. A*, 88:032111, Sep 2013.

[13] Scott Aaronson, Adam Bouland, Joseph Fitzsimons, and Mitchell Lee. The space "just above" BQP. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, Cambridge, MA, USA, January 14-16, 2016*, pages 271–280, 2016.

[14] Scott Aaronson, Adam Bouland, Greg Kuperberg, and Saeed Mehraban. The computational complexity of ball permutations. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 317–327. ACM, 2017.

[15] Scott Aaronson and Lijie Chen. Complexity-theoretic foundations of quantum supremacy experiments. *arXiv preprint arXiv:1612.05903*, 2016.

[16] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004.

[17] Scott Aaronson, Daniel Grier, and Luke Schaeffer. The classification of reversible bit operations. In *Proceedings of Innovations in Theoretical Computer Science (ITCS)*, 2017.

[18] Scott Aaronson, Greg Kuperberg, and Christopher Granade. The complexity zoo, 2005.

[19] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *Journal of the ACM (JACM)*, 51(4):595–605, 2004.

[20] Scott Aaronson and John Watrous. Closed timelike curves make quantum and classical computing equivalent. *Proceedings of the Royal Society A*, 465:631–647, 2009.

[21] Scott Aaronson and Avi Wigderson. Algebrization: A new barrier in complexity theory. *ACM Transactions on Computation Theory (TOCT)*, 1(1):2, 2009.

[22] Daniel S. Abrams and Seth Lloyd. Nonlinear quantum mechanics implies polynomial-time solution for NP-complete and #P problems. *Physical Review Letters*, 81:3992, 1998. arXiv:quant-ph/9801041.

[23] Samson Abramsky and Bob Coecke. Categorical quantum mechanics. In K. Engesser, D. Gabbay, and D. Lehmann, editors, *Handbook of Quantum Logic and Quantum Structures*, volume 2, pages 261–325. Elsevier, 2008. arXiv:0808.1023.

[24] Leonard M Adleman, Jonathan DeMarrais, and Ming-Deh A Huang. Quantum computability. *SIAM Journal on Computing*, 26(5):1524–1540, 1997.

[25] Stephen L. Adler. *Quaternionic Quantum Mechanics and Quantum Fields*. Oxford University Press, Oxford, 1995.

[26] Shweta Agrawal, Yuval Ishai, Dakshita Khurana, and Anat Paskin-Cherniavsky. Statistical randomized encodings: A complexity theoretic view. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 1–13, 2015.

[27] Dorit Aharonov. A simple proof that Toffoli and Hadamard are quantum universal. 2003. arXiv:quant-ph/0301040.

[28] Dorit Aharonov and Itai Arad. The BQP-hardness of approximating the Jones polynomial. *New Journal of Physics*, 13(3):035019, 2011. arXiv:quant-ph/0605181.

[29] Dorit Aharonov, Itai Arad, Elad Eban, and Zeph Landau. Polynomial quantum algorithms for additive approximations of the potts model and other points of the tutte plane. *arXiv preprint quant-ph/0702008*, 2007.

[30] Dorit Aharonov and Michael Ben-Or. Fault-tolerant quantum computation with constant error. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 176–188. ACM, 1997.

[31] Dorit Aharonov, Vaughan Jones, and Zeph Landau. A polynomial quantum algorithm for approximating the Jones polynomial. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 427–436. ACM, 2006.

[32] Dorit Aharonov, Wim Van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM Journal on Computing*, 37(1):166–194, 2007.

[33] William Aiello and Johan Håstad. Relativized perfect zero knowledge is not BPP. *Information and Computation*, 93:223–240, 1991.

[34] William Aiello and Johan Håstad. Statistical zero-knowledge languages can be recognized in two rounds. *J. Comput. Syst. Sci.*, 42(3):327–345, 1991.

[35] Ahmed Almheiri, Donald Marolf, Joseph Polchinski, and James Sully. Black holes: Complementarity or firewalls? *Journal of High Energy Physics*, 1302:062, 2013. arXiv:1207.3123.

[36] Andris Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.

[37] Andris Ambainis, Leonard J Schulman, and Umesh V Vazirani. Computing with highly mixed states. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 697–704. ACM, 2000.

[38] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[39] L Babai. Trading group theory for randomness. In *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*, STOC '85, pages 421–429, New York, NY, USA, 1985. ACM.

[40] László Babai. Graph isomorphism in quasipolynomial time [extended abstract]. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 684–697. ACM, 2016.

[41] D Bacon, J Kempe, DP DiVincenzo, DA Lidar, and KB Whaley. Encoded universality in physical implementations of a quantum computer. *arXiv:quant-ph/0102140*, 2001.

[42] D Bacon, J Kempe, DA Lidar, and KB Whaley. Universal fault-tolerant computation on decoherence-free subspaces. *Physical Review Letters*, 85(8):1758–61, 2000. arXiv:quant-ph/9909058.

[43] Dave Bacon. Quantum computational complexity in the presence of closed timelike curves. *Physical Review A*, 70(3):032309, 2004.

[44] J. Bae, W-Y. Hwang, and Y-D. Han. No-signaling principle can determine optimal quantum state discrimination. *Physical Review Letters*, 107:170403, 2011. arXiv:1102.0361.

[45] Theodore Baker, John Gill, and Robert Solovay. Relativizations of the P=?NP question. *SIAM Journal on computing*, 4(4):431–442, 1975.

[46] Ning Bao, Adam Bouland, and Stephen P. Jordan. Grover search and the no-signaling principle. *Phys. Rev. Lett.*, 117:120501, Sep 2016.

[47] Ning Bao, Adam Bouland, Jason Pollack, Henry Yuen, and Aidan Chatwin-Davies. Rescuing complementarity with little drama. *JHEP*, 1612(arXiv: 1607.05141):026, 2016.

[48] Boaz Barak. How to go beyond the black-box simulation barrier. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 106–115, 2001.

[49] Adriano Barenco, Charles H Bennett, Richard Cleve, David P DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A Smolin, and Harald Weinfurter. Elementary gates for quantum computation. *Physical review A*, 52(5):3457, 1995.

[50] Jonathan Barrett. Information processing in generalized probabilistic theories. *Physical Review A*, 75:032304, 2005. arXiv:quant-ph/0508211.

[51] Stephen D Bartlett and Barry C Sanders. Requirement for quantum computation. *Journal of Modern Optics*, 50(15-17):2331–2340, 2003.

[52] Bela Bauer, Claire Levaillant, and Michael Freedman. Universality of single quantum gates. *arXiv preprint arXiv:1404.7822*, 2014.

[53] Alan F Beardon. *The geometry of discrete groups*, volume 91. Springer Science & Business Media, 2012.

[54] Richard Beigel, Nick Reingold, and Daniel A. Spielman. PP is closed under intersection (extended abstract). In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, May 5-8, 1991, New Orleans, Louisiana, USA*, pages 1–9, 1991.

[55] Jacob D Bekenstein. Black holes and entropy. *Physical Review D*, 7(8):2333, 1973.

[56] Charles H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997.

[57] Charles H. Bennett, Gilles Brassard, Claude Crepau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895, 1993.

[58] Charles H. Bennett, Debbie Leung, Graeme Smith, and John A. Smolin. Can closed timelike curves or nonlinear quantum mechanics improve quantum state discrimination or help solve hard problems? *Physical Review Letters*, 103:170502, 2009. arXiv:0908.3023.

[59] Charles H. Bennett, Debbie Leung, Graeme Smith, and John A. Smolin. Can closed timelike curves or nonlinear quantum mechanics improve quantum state discrimination or help solve hard problems? *Physical Review Letters*, 103:170502, 2009.

[60] Alex Bocharov, Martin Roetteler, and Krysta M Svore. Efficient synthesis of universal repeat-until-success quantum circuits. *Physical review letters*, 114(8):080502, 2015.

[61] Sergio Boixo, Sergei V Isakov, Vadim N Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, John M Martinis, and Hartmut Neven. Characterizing quantum supremacy in near-term devices. *arXiv preprint arXiv:1608.00263*, 2016.

[62] Ravi B. Boppana, Johan Håstad, and Stathis Zachos. Does co-NP have short interactive proofs? *Inf. Process. Lett.*, 25(2):127–132, 1987.

[63] Adam Bouland and Scott Aaronson. Generation of universal linear optics by any beam splitter. *Physical Review A*, 89(6):062316, 2014.

[64] Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan. On the power of statistical zero knowledge. *arXiv:1609.02888*, 2016. To appear in Proc. FOCS 2017.

[65] Adam Bouland, Joseph Fitzsimons, and Dax Koh. Quantum advantage from conjugated Clifford circuits. *In prep.*, 2017.

[66] Adam Bouland, Laura Mančinska, and Xue Zhang. Complexity Classification of Two-Qubit Commuting Hamiltonians. In Ran Raz, editor, *31st Conference on Computational Complexity (CCC 2016)*, volume 50 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 28:1–28:33, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[67] Raphael Bousso. Complementarity is not enough. *Physical Review D*, 87(12):124023, 2013.

[68] Samuel L. Braunstein, Stefano Pirandola, and Karol Życzkowski. Better late than never: Information retrieval from black holes. *Physical Review Letters*, 110:101301, 2013.

[69] Sergey Bravyi and Jeongwan Haah. Magic-state distillation with low overhead. *Physical Review A*, 86(5):052329, 2012.

[70] Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal Clifford gates and noisy ancillas. *Physical Review A*, 71(2):022316, 2005.

[71] Michael J Bremner, Christopher M Dawson, Jennifer L Dodd, Alexei Gilchrist, Aram W Harrow, Duncan Mortimer, Michael A Nielsen, and Tobias J Osborne. Practical scheme for quantum computation with any two-qubit entangling gate. *Physical review letters*, 89(24):247902, 2002.

[72] Michael J Bremner, Richard Jozsa, and Dan J Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, page rspa20100301. The Royal Society, 2010.

[73] Michael J Bremner, Ashley Montanaro, and Dan J Shepherd. Achieving quantum supremacy with sparse and noisy commuting quantum computations. *arXiv preprint arXiv:1610.01808*, 2016.

[74] Michael J Bremner, Ashley Montanaro, and Dan J Shepherd. Average-case complexity versus approximate simulation of commuting quantum computations. *Physical review letters*, 117(8):080501, 2016.

[75] Hans J Briegel, David E Browne, W Dür, Robert Raussendorf, and Maarten Van den Nest. Measurement-based quantum computation. *Nature Physics*, 5(1):19–26, 2009.

[76] Theodor Bröcker and Tammo tom Dieck. *Representations of compact Lie groups*, volume 98. Springer Science & Business Media, 2013.

[77] Daniel J Brod and Andrew M Childs. The computational power of matchgates and the XY interaction on arbitrary graphs. *Quantum Information & Computation*, 14(11-12):901–916, 2014.

[78] Todd Brun. Computers with closed timelike curves can solve hard problems. *Foundations of Physics Letters*, 16:245–253, 2003. arXiv:gr-qc/0209061.

[79] Jean-Luc Brylinski and Ranee Brylinski. Universal quantum gates. *Mathematics of Quantum Computation*, 79, 2002.

[80] Mark Bun and Justin Thaler. Dual polynomials for collision and element distinctness. *arXiv preprint arXiv:1503.07261*, 2015.

[81] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *International Colloquium on Automata, Languages, and Programming*, pages 268–280. Springer, 2015.

[82] Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of AC^0. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:51, 2017.

[83] Elie Cartan. La théorie des groupes finis et continus et l'analysis situs. *Mémorial des sciences mathématiques*, 42:1–61, 1952.

[84] Stefania Cavallar, Bruce Dodson, Arjen Lenstra, Walter Lioen, Peter Montgomery, Brian Murphy, Herman Te Riele, Karen Aardal, Jeff Gilchrist, Gérard Guillerm, et al. Factorization of a 512-bit RSA modulus. In *Advances in Cryptology—EUROCRYPT 2000*, pages 1–18. Springer, 2000.

[85] Nicolas J Cerf, Christoph Adami, and Paul G Kwiat. Optical simulation of quantum logic. *Physical Review A*, 57(3):R1477, 1998.

[86] Amit Chakrabarti, Graham Cormode, Andrew McGregor, Justin Thaler, and Suresh Venkatasubramanian. Verifiable stream computation and Arthur-Merlin communication. In David Zuckerman, editor, *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, volume 33 of *LIPIcs*, pages 217–243. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015.

[87] Richard Chang, Benny Chor, Oded Goldreich, Juris Hartmanis, Johan Håstad, Desh Ranjan, and Pankaj Rohatgi. The random oracle hypothesis is false. *J. Comput. Syst. Sci.*, 49(1):24–39, 1994.

[88] Lijie Chen. Adaptivity vs postselection. *arXiv:1606.04016*, 2016.

[89] Lijie Chen. A note on oracle separations for BQP. *arXiv:1605.00619*, 2016.

[90] Andrew M Childs, Debbie Leung, Laura Mančinska, and Māris Ozols. Characterization of universal two-qubit Hamiltonians. *Quantum Information & Computation*, 11(1):19–39, 2011.

[91] Andrew M. Childs and Joshua Young. Optimal state discrimination and unstructured search in nonlinear quantum mechanics. 2015. arXiv:1507.06334.

[92] Richard Cleve and John Watrous. Fast parallel circuits for the quantum Fourier transform. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 526–536. IEEE, 2000.

[93] Benoit Collins and Piotr Sniady. Integration with respect to the Haar measure on unitary, orthogonal, and symplectic group. *Communications in Mathematical Physics*, 264:773–795, 2006.

[94] Graham Cormode, Justin Thaler, and Ke Yi. Verifying computations with streaming interactive proofs. *PVLDB*, 5(1):25–36, 2011.

[95] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley, Hoboken, New Jersey, 1991.

[96] W.J.R. Crosby. Solution of the problem 4136. *Amer. Math. Monthly*, 53:103–107, 1946.

[97] Toby Cubitt and Ashley Montanaro. Complexity classification of local Hamiltonian problems. *SIAM Journal on Computing*, 45(2):268–316, 2016.

[98] Ivan Damgård, Oded Goldreich, Tatsuaki Okamoto, and Avi Wigderson. Honest verifier vs dishonest verifier in public coin zero-knowledge proofs. In *Advances in Cryptology - CRYPTO '95, 15th Annual International Cryptology Conference, Santa Barbara, California, USA, August 27-31, 1995, Proceedings*, pages 325–338, 1995.

[99] Christopher M Dawson and Michael A Nielsen. The Solovay-Kitaev algorithm. *Quantum Information & Computation*, 6(1):81–95, 2006.

[100] David Deutsch. Quantum theory, the Church-Turing principle and the universal quantum computer. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 400, pages 97–117. The Royal Society, 1985.

[101] David Deutsch. Quantum mechanics near closed timelike lines. *Physical Review D*, 44:3197–3217, 1991.

[102] David Deutsch, Adriano Barenco, and Artur Ekert. Universality in quantum computation. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 449, pages 669–677. The Royal Society, 1995.

[103] David P DiVincenzo, Dave Bacon, Julia Kempe, Guido Burkard, and K Birgitta Whaley. Universal quantum computation with the exchange interaction. *Nature*, 408(6810):339–342, 2000.

[104] David P. DiVincenzo and Peter W. Shor. Fault-tolerant error correction with efficient quantum codes. *Phys. Rev. Lett.*, 77:3260–3263, Oct 1996.

[105] Jennifer L Dodd, Michael A Nielsen, Michael J Bremner, and Robert T Thew. Universal quantum computation and simulation using any entangling Hamiltonian and local unitaries. *Physical Review A*, 65(4):040301, 2002.

[106] Bryan Eastin and Emanuel Knill. Restrictions on transversal encoded quantum gate sets. *Physical review letters*, 102(11):110502, 2009.

[107] Lior Eldar and Peter Shor. A discrete fourier transform on lattices with quantum applications. *arXiv preprint arXiv:1703.02515*, 2017.

[108] Mark Ettinger, Peter Høyer, and Emanuel Knill. The quantum query complexity of the hidden subgroup problem is polynomial. *Information Processing Letters*, 91(1):43–48, 2004.

[109] W. M. Fairbairn, T. Fulton, and W. H. Klink. Finite and Disconnected Subgroups of SU3 and their Application to the Elementary-Particle Spectrum. *Journal of Mathematical Physics*, 5(8):1038, 1964.

[110] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution. *arXiv preprint quant-ph/0001106*, 2000.

[111] Edward Farhi and Aram W Harrow. Quantum supremacy through the quantum approximate optimization algorithm. *arXiv preprint arXiv:1602.07674*, 2016.

[112] Bill Fefferman, Michael Foss-Feig, and Alexey V Gorshkov. Exact sampling hardness of Ising spin models. *arXiv:1701.03167*, 2017.

[113] Bill Fefferman and Christopher Umans. Pseudorandom generators and the BQP vs. PH problem. *arXiv preprint arXiv:1007.0305*, 2010.

[114] Bill Fefferman and Christopher Umans. On the power of quantum Fourier sampling. In *11th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2016, September 27-29, 2016, Berlin, Germany*, pages 1:1–1:19, 2016.

[115] Richard P Feynman. Simulating physics with computers. *International journal of theoretical physics*, 21(6):467–488, 1982.

[116] Marc Fischlin. On the impossibility of constructing non-interactive statistically-secret protocols from any trapdoor one-way function. In *Proceedings of the The Cryptographer's Track at the RSA Conference on Topics in Cryptology*, CT-RSA '02, pages 79–95, London, UK, UK, 2002. Springer-Verlag.

[117] Bryan H Fong and Stephen M Wandzura. Universal quantum computation and leakage reduction in the 3-qubit decoherence free subsystem. *Quantum Information & Computation*, 11(11-12):1003–1018, 2011. arXiv:1102.2909.

[118] Lance Fortnow. The complexity of perfect zero-knowledge (extended abstract). In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 1987, New York, New York, USA*, pages 204–209, 1987.

[119] Lance Fortnow and John D. Rogers. Complexity limitations on quantum computation. *J. Comput. Syst. Sci.*, 59(2):240–252, 1999. arXiv:cs/9811023.

[120] Michael Freedman, Alexei Kitaev, Michael Larsen, and Zhenghan Wang. Topological quantum computation. *Bulletin of the American Mathematical Society*, 40(1):31–38, 2003. arXiv:quant-ph/0101025.

[121] Michael H. Freedman, Alexei Kitaev, and Zhenghan Wang. Simulation of topological field theories by quantum computers. *Commun. Math. Phys.*, 227:587–603, 2002.

[122] Keisuke Fujii, Hirotada Kobayashi, Tomoyuki Morimae, Harumichi Nishimura, Shuhei Tamate, and Seiichiro Tani. Impossibility of classically simulating one-clean-qubit computation. *arXiv preprint arXiv:1409.6777*, 2014.

[123] Craig Gidney. Factoring with n+2 clean qubits and n-1 dirty qubits. *arXiv:1706.07884*, 2017.

[124] Nicolas Gisin. Weinberg's non-linear quantum mechanics and supraluminal communications. *Physics Letters A*, 143(1-2):1–2, 1990.

[125] Oded Goldreich and Shafi Goldwasser. On the limits of non-approximability of lattice problems. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 1–9. ACM, 1998.

[126] Oded Goldreich, Silvio Micali, and Avi Wigderson. Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems. *Journal of the ACM (JACM)*, 38(3):690–728, 1991.

[127] Oded Goldreich, Amit Sahai, and Salil Vadhan. Can statistical zero knowledge be made non-interactive? Or on the relationship of SZK and NISZK. In *Advances in Cryptology—CRYPTO'99*, pages 467–484. Springer, 1999.

[128] Oded Goldreich, Amit Sahai, and Salil P. Vadhan. Honest-verifier statistical zero-knowledge equals general statistical zero-knowledge. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 399–408, 1998.

[129] Oded Goldreich and Liav Teichner. Super-perfect zero-knowledge proofs. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:97, 2014.

[130] Shafi Goldwasser. Zero knowledge probabilistic proof systems, 2015.

[131] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on computing*, 18(1):186–208, 1989.

[132] Daniel Gottesman. Stabilizer codes and quantum error correction. *arXiv preprint quant-ph/9705052*, 1997.

[133] Daniel Gottesman. The Heisenberg representation of quantum computers. *Group22: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics*, pages 32–43, 1999.

[134] Daniel Gottesman and Isaac L Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402(6760):390–393, 1999.

[135] Daniel Gottesman, Alexei Kitaev, and John Preskill. Encoding a qubit in an oscillator. *Physical Review A*, 64(1):012310, 2001.

[136] Daniel Gottesman and John Preskill. Comment on "The black hole final state". *Journal of High Energy Physics*, 0403:026, 2004. arXiv:hep-th/0311269.

[137] Daniel Grier and Luke Schaeffer. The classification of stabilizer operations over qubits. *arXiv preprint arXiv:1603.03999*, 2016.

[138] Walter Grimus and Patrick Otto Ludl. Finite flavour groups of fermions. *Journal of Physics A: Mathematical and Theoretical*, 45(23):233001, 2012.

[139] Walter Grimus and Patrick Otto Ludl. On the characterization of the SU(3)-subgroups of type C and D. *Journal of Physics A: Mathematical and Theoretical*, 47(7):075202, 2014.

[140] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 212–219, 1996. arXiv:quant-ph/9605043.

[141] Brian C Hall. *Lie groups, Lie algebras, and representations: an elementary introduction*, volume 222. Springer, 2015.

[142] Amihay Hanany and Yang-Hui He. A monograph on the classification of the discrete subgroups of SU(4). *Journal of High Energy Physics*, 2001(02):027, 2001.

[143] Amihay Hanany and YH He. Non-abelian finite gauge theories. *Journal of High Energy Physics*, 1999(02):013, 1999.

[144] Thomas Häner, Martin Roetteler, and Krysta M Svore. Factoring using 2n+ 2 qubits with Toffoli based modular multiplication. *arXiv preprint arXiv:1611.07995*, 2016.

[145] Dominik Hangleiter, Juan Bermejo-Vega, Martin Schwarz, and Jens Eisert. Anti-concentration theorems for schemes showing a quantum computational supremacy. *arXiv:1706.03786*, 2017.

[146] Daniel Harlow. Personal communication.

[147] Daniel Harlow. Aspects of the Papadodimas-Raju proposal for the black hole interior. *Journal of High Energy Physics*, page 1411, 2014. arXiv:1405.1995.

[148] Daniel Harlow. Jerusalem lectures on black holes and quantum information. 2014. arXiv:1409.1231.

[149] Daniel Harlow. Jerusalem lectures on black holes and quantum information. *Reviews of Modern Physics*, 88(1):015002, 2016.

[150] Daniel Harlow and Patrick Hayden. Quantum computation vs. firewalls. *Journal of High Energy Physics*, 1013:85, 2013. arXiv:1301.4504.

[151] Aram Harrow and Saeed Mehraban. *Personal Communication*, 2017.

[152] S. W. Hawking. Breakdown of predictability in gravitational collapse. *Phys. Rev. D*, 14:2460–2473, Nov 1976.

[153] Stephen W Hawking. Particle creation by black holes. *Communications in mathematical physics*, 43(3):199–220, 1975.

[154] Thomas Holenstein and Renato Renner. One-way secret-key agreement and applications to circuit polarization and immunization of public-key encryption. In *Advances in Cryptology - CRYPTO 2005: 25th Annual International Cryptology Conference, Santa Barbara, California, USA, August 14-18, 2005, Proceedings*, pages 478–493, 2005.

[155] Gary T. Horowitz and Juan Maldecena. The black hole final state. *Journal of High Energy Physics*, 0402:008, 2004. arXiv:hep-th/0310281.

[156] Russell Impagliazzo and Avi Wigderson. P= BPP if E requires exponential circuits: Derandomizing the XOR lemma. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 220–229. ACM, 1997.

[157] Gábor Ivanyos. Deciding universality of quantum gates. *Journal of Algebra*, 310(1):49–56, 2007.

[158] Gordon James and Adalbert Kerber. The representation theory of the symmetric group. *Reading, Mass*, 1981.

[159] Stephen P Jordan. Permutational quantum computing. *arXiv:0906.2508*, 2009.

[160] Stephen P. Jordan, Keith S. M. Lee, and John Preskill. Quantum algorithms for quantum field theories. *Science*, 336(6085):1130–1133, 2012.

[161] Stephen P. Jordan, Keith S. M. Lee, and John Preskill. Quantum algorithms for fermionic quantum field theories. 2014. arXiv:1404.7115.

[162] Stephen P. Jordan, Keith S. M. Lee, and John Preskill. Quantum computation of scattering in scalar quantum field theories. *Quantum Information and Computation*, 14:1014–1080, 2014.

[163] Richard Jozsa and Maarten Van den Nest. Classical simulation complexity of extended Clifford circuits. *Quantum Information and Computation*, 14(7/8):633–648, 2014.

[164] Varun Kanade and Justin Thaler. Distribution-independent reliable learning. In Maria-Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, volume 35 of *JMLR Workshop and Conference Proceedings*, pages 3–24. JMLR.org, 2014.

[165] Phillip Kaye, Raymond Laflamme, and Michele Mosca. *An Introduction to Quantum Computing*. Oxford University Press, Oxford, 2007.

[166] Julia Kempe, Dave Bacon, Daniel A Lidar, and K Birgitta Whaley. Theory of decoherence-free fault-tolerant universal quantum computation. *Physical Review A*, 63(4):042307, 2001. arXiv:quant-ph/0004064.

[167] Julia Kempe, David Bacon, David P DiVincenzo, and K Brigitta Whaley. Encoded universality from a single physical interaction. *Quantum Information & Computation*, 1(4):33–55, 2001. arXiv:quant-ph/0112013.

[168] Julia Kempe and K Birgitta Whaley. Exact gate sequences for universal quantum computation using the XY interaction alone. *Physical Review A*, 65(5):052330, 2002. arXiv:quant-ph/0112014.

[169] A Yu Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[170] K. v. Klitzing, G. Dorda, and M. Pepper. New method for high-accuracy determination of the fine-structure constant based on quantized hall resistance. *Phys. Rev. Lett.*, 45:494–497, Aug 1980.

[171] Emanuel Knill and Raymond Laflamme. Power of one bit of quantum information. *Physical Review Letters*, 81(25):5672, 1998.

[172] Emanuel Knill, Raymond Laflamme, and Gerald J Milburn. A scheme for efficient quantum computation with linear optics. *nature*, 409(6816):46–52, 2001.

[173] Dax Enshan Koh. Further extensions of Clifford circuits and their classical simulation complexities. *arXiv preprint arXiv:1512.07892*, 2015.

[174] Greg Kuperberg. The capacity of hybrid quantum memory. *Information Theory, IEEE Transactions on*, 49(6):1465–1473, 2003. arXiv:quant-ph/0203105.

[175] Greg Kuperberg. Denseness and Zariski denseness of Jones braid representations. *Geometry & Topology*, 15(1):11–39, 2011. arXiv:0909.1881.

[176] Greg Kuperberg. How hard is it to approximate the Jones polynomial? *Theory of Computing*, 11(6):183–219, 2015.

[177] Samuel Kutin. Quantum lower bound for the collision problem with small range. *Theory of Computing*, 1(1):29–36, 2005.

[178] Raymond Laflamme, Cesar Miquel, Juan Pablo Paz, and Wojciech Hubert Zurek. Perfect quantum error correcting code. *Physical Review Letters*, 77(1):198, 1996.

[179] Martin Leslie, 2012.

[180] Seth Lloyd. Almost any quantum logic gate is universal. *Physical Review Letters*, 75(2):346, 1995.

[181] Seth Lloyd and John Preskill. Unitarity of black hole evaporation in final-state projection models. *Journal of High Energy Physics*, 08:126, 2014. arXiv:1308.4209.

[182] Shachar Lovett and Jiapeng Zhang. On the impossibility of entropy reversal, and its application to zero-knowledge proofs. *ECCC TR16-118*, July 31 2016.

[183] Patrick Otto Ludl. Comments on the classification of the finite subgroups of su (3). *Journal of Physics A: Mathematical and Theoretical*, 44(25):255204, 2011.

[184] Lior Malka. How to achieve perfect simulation and a complete problem for non-interactive perfect zero-knowledge. *Journal of Cryptology*, 28(3):533–550, 2015.

[185] Ryan L. Mann and Michael J. Bremner. *Personal Communication*, 2017.

[186] Donald Marolf and Joseph Polchinski. Violations of the Born rule in cool state-dependent horizons. 2015. arXiv:1506.01337.

[187] Enrique Martin-Lopez, Anthony Laing, Thomas Lawson, Roberto Alvarez, Xiao-Qi Zhou, and Jeremy L O'Brien. Experimental realization of shor's quantum factoring algorithm using qubit recycling. *Nature Photonics*, 6(11):773–776, 2012.

[188] Dmitri Maslov and Martin Roetteler. Shorter stabilizer circuits via Bruhat decomposition and quantum circuit transformations. *arXiv preprint arXiv:1705.09176*, 2017.

[189] David A. Meyer and Thomas G. Wong. Nonlinear quantum search using the Gross-Pitaevskii equation. *New Journal of Physics*, 15:063014, 2013. arXiv:1303.0371.

[190] David A. Meyer and Thomas G. Wong. Quantum search with general nonlinearities. *Physical Review A*, 89:012312, 2014. arXiv:1310.7301.

[191] David A. Meyer and Thomas G. Wong. Completeness is unnecessary for fast nonlinear quantum search. 2015. arXiv:1502.06281.

[192] Peter Bro Miltersen and N Variyam Vinodchandran. Derandomizing Arthur-Merlin games using hitting sets. In *Foundations of Computer Science, 1999. 40th Annual Symposium on*, pages 71–80. IEEE, 1999.

[193] Masoud Mohseni, Peter Read, Hartmut Neven, Sergio Boixo, Vasil Denchev, Ryan Babbush, Austin Fowler, Vadim Smelyanskiy, and John Martinis. Commercialize quantum technologies in five years. *Nature*, 543:171–174, 2017.

[194] Tomoyuki Morimae. Hardness of classically sampling one clean qubit model with constant total variation distance error. *arXiv preprint arXiv:1704.03640*, 2017.

[195] Tomoyuki Morimae, Keisuke Fujii, and Joseph F Fitzsimons. Hardness of classically simulating the one-clean-qubit model. *Physical review letters*, 112(13):130502, 2014.

[196] Gabriele Nebe, Eric M Rains, and Neil JA Sloane. The invariants of the Clifford groups. *Designs, Codes and Cryptography*, 24(1):99–122, 2001.

[197] Gabriele Nebe, Eric M Rains, and Neil James Alexander Sloane. *Self-dual codes and invariant theory*, volume 17. Springer, 2006.

[198] Xiaotong Ni and Maarten Van Den Nest. Commuting quantum circuits: efficient classical simulations versus hardness results. *Quantum Information & Computation*, 13(1-2):54–72, 2013.

[199] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[200] Noam Nisan and Avi Wigderson. Hardness vs randomness. *Journal of computer and System Sciences*, 49(2):149–167, 1994.

[201] Tatsuaki Okamoto. On relationships between statistical zero-knowledge proofs. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 649–658. ACM, 1996.

[202] Michał Oszmaniec and Zoltán Zimborás. Universal extensions of restricted classes of quantum operations. *arXiv preprint arXiv:1705.11188*, 2017.

[203] K. Papadodimas and S. Raju. State-dependent bulk-boundary maps and black hole complementarity. *Physical Review D*, 89:086010, 2014. arXiv:1310.6335.

[204] Chris Peikert and Vinod Vaikuntanathan. Noninteractive statistical zero-knowledge proofs for lattice problems. In *Annual International Cryptology Conference*, pages 536–553. Springer, 2008.

[205] Asher Peres. Two simple proofs of the Kochen-Specker theorem. *Journal of Physics A: Mathematical and General*, 24(4):L175, 1991.

[206] Joseph Polchinski. Weinberg's nonlinear quantum mechanics and the Einstein-Podolsky-Rosen paradox. *Physical Review Letters*, 66:397, 1991.

[207] John Preskill. Quantum computing and the entanglement frontier. *arXiv preprint arXiv:1203.5813*, 2012.

[208] Michael Reck, Anton Zeilinger, Herbert J Bernstein, and Philip Bertani. Experimental realization of any discrete unitary operator. *Physical Review Letters*, 73(1):58, 1994.

[209] Neil J Ross and Peter Selinger. Optimal ancilla-free Clifford+ T approximation of z-rotations. *arXiv preprint arXiv:1403.2975*, 2014.

[210] Remo Ruffini and Silvano Bonazzola. Systems of self-gravitating particles in general relativity and the concept of an equation of state. *Physical Review*, 187:1767, 1969.

[211] Amit Sahai and Salil Vadhan. A complete problem for statistical zero knowledge. *Journal of the ACM (JACM)*, 50(2):196–249, 2003.

[212] Imdad SB Sardharwalla, Toby S Cubitt, Aram W Harrow, and Noah Linden. Universal refocusing of systematic quantum noise. *arXiv preprint arXiv:1602.07963*, 2016.

[213] Adam Sawicki. Universality of beamsplitters. *arXiv preprint arXiv:1507.08255*, 2015.

[214] Adam Sawicki and Katarzyna Karnas. Criteria for universality of quantum gates. *arXiv preprint arXiv:1610.00547*, 2016.

[215] Dan Shepherd. Binary matroids and quantum probability distributions. *arXiv preprint arXiv:1005.1744*, 2010.

[216] Dan Shepherd and Michael J Bremner. Temporally unstructured quantum computation. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 465, pages 1413–1439. The Royal Society, 2009.

[217] Daniel James Shepherd. Quantum complexity: restrictions on algorithms and architectures. *PhD Thesis, University of Bristol*, 2009.

[218] Alexander A Sherstov. Breaking the Minsky-Papert barrier for constant-depth circuits. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 223–232. ACM, 2014.

[219] Alexander A Sherstov. The power of asymmetry in constant-depth circuits. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 431–450. IEEE, 2015.

[220] Yaoyun Shi. Both Toffoli and controlled-NOT need little help to do universal quantum computing. *Quantum Information & Computation*, 3(1):84–92, 2003.

[221] Peter W Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303–332, 1999.

[222] Peter W Shor and Stephen P Jordan. Estimating Jones polynomials is a complete problem for one clean qubit. *Quantum Information & Computation*, 8(8):681–714, 2008. arXiv:0707.2831.

[223] John A Smolin, Graeme Smith, and Alexander Vargo. Oversimplifying quantum factoring. *Nature*, 499(7457):163–165, 2013.

[224] Andrew Steane. Multiple-particle interference and quantum error correction. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 452, pages 2551–2577. The Royal Society, 1996.

[225] Larry Stockmeyer. The complexity of approximate counting. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 118–126. ACM, 1983.

[226] Leonard Susskind. *The black hole war: My battle with Stephen Hawking to make the world safe for quantum mechanics.* Hachette UK, 2008.

[227] Leonard Susskind, Larus Thorlacius, and John Uglum. The stretched horizon and black hole complementarity. *Physical Review D*, 48(8):3743, 1993.

[228] Yasuhiro Takahashi, Seiichiro Tani, Takeshi Yamazaki, and Kazuyuki Tanaka. Commuting quantum circuits with few outputs are unlikely to be classically simulatable. In *International Computing and Combinatorics Conference*, pages 223–234. Springer, 2015.

[229] Barbara M Terhal and David P DiVincenzo. Classical simulation of noninteracting-fermion quantum circuits. *Physical Review A*, 65(3):032325, 2002.

[230] Barbara M Terhal and David P DiVincenzo. Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games. *Quantum Information & Computation*, 4(2):134–145, 2004.

[231] Justin Thaler. Lower bounds for the approximate degree of block-composed functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:150, 2014.

[232] Seinosuke Toda. PP is as hard as the polynomial-time hierarchy. *SIAM J. Comput.*, 20(5):865–877, 1991.

[233] Tommaso Toffoli. Reversible computing. *Automata, languages and programming*, pages 632–644, 1980.

[234] Gregory Valiant and Paul Valiant. Estimating the unseen: an n/log(n)-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 685–694, 2011.

[235] Leslie Valiant and Umesh Vazirani. NP is as easy as detecting unique solutions. *Theoretical Computer Science*, 47:85–93, 1986.

[236] Leslie G Valiant. Quantum computers that can be simulated classically in polynomial time. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 114–123. ACM, 2001.

[237] Lieven MK Vandersypen, Matthias Steffen, Gregory Breyta, Costantino S Yannoni, Mark H Sherwood, and Isaac L Chuang. Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature*, 414(6866):883–887, 2001.

[238] Nikolai K. Vereshchagin. Lower bounds for perceptrons solving some separation problems and oracle separation of AM from PP. In *Third Israel Symposium on Theory of Computing and Systems, ISTCS 1995, Tel Aviv, Israel, January 4-6, 1995, Proceedings*, pages 46–51, 1995.

[239] John Watrous. Succinct quantum proofs for properties of finite groups. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 537–546. IEEE, 2000.

[240] Nik Weaver. On the universality of almost every quantum logic gate. *Journal of Mathematical Physics*, 41(1):240–243, 2000.

[241] Zak Webb. The Clifford group forms a unitary 3-design. *Quantum Information and Computation*, 16:1379–1400, 2016.

[242] Eugene P. Wigner. *Gruppentheorie und ihre Anwendung auf die Quanten mechanik der Atomspektren*, pages 251–254. Friedrich Vieweg und Sohn, 1931.

[243] L. Włodarski. On the equation $\cos(\alpha_1) + \cos(\alpha_2) + \cos(\alpha_3) + \cos(\alpha_4) = 0$. *Ann. Univ. Sci. Budapest. Eötvös Sect. Math*, 12:147, 1969.

[244] L-A Wu and DA Lidar. Power of anisotropic exchange interactions: Universality and efficient codes for quantum computing. *Physical Review A*, 65(4):042318, 2002.

[245] Q. Yuan. The representation theory of SU(2). *Blog entry from "Annoying Precision". http://qchu.wordpress.com/2011/06/26/the-representation-theory-of-su2/ Retrieved June 16, 2017*, 2011.

[246] Christof Zalka. Grover's quantum searching algorithm is optimal. *Physical Review A*, 60:2746–2751, 1999. arXiv:quant-ph/9711070.

[247] Huangjun Zhu. Multiqubit Clifford groups are unitary 3-designs. *arXiv:1510.02619.*