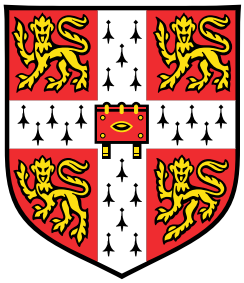
The power of restricted quantum computational models



Mithuna Yoganathan

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

January 2021

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified here.

This thesis is based on three papers. The first is "Quantum advantage of unitary Clifford circuits with magic state inputs" by Mithuna Yoganathan, Richard Jozsa, and Sergii Strelchuk, published in The Proceedings of the Royal Society [92]. The author contributed the idea for the project and the most of the technical results.

The second paper is "The one clean qubit model without entanglement is classically simulable" by Mithuna Yoganathan and Chris Cade [91]. This paper was accepted for a talk at the 2020 Quantum Information Processing conference, and is to be submitted to Physical Review Letters. The author contributed the idea for the project and the most of the technical results.

The final paper is "A condition under which classical simulability implies efficient state learnability" by Mithuna Yoganathan. This paper is being prepared for publication.

Mithuna Yoganathan
January 2021

Abstract

Restricted models of quantum computation are ones that have less power than a universal quantum computer. We studied the consequences of removing particular properties from a universal quantum computer to discover whether those resources were important.

In the first part of the thesis we studied universal quantum computers which are implemented using Clifford gates, adaptive measurements, and magic states. The Gottesman–Knill theorem shows that circuits in this form which do not use magic states can be simulated by a classical computer. We extended this result to show that *all* circuits in this form can be partially simulated; the same computation can be implemented using a smaller quantum computer with the assistance of some polynomial time classical computation. We also identified a subclass of these computations that can be shown to not be entirely classically simulated by any method, given certain complexity theoretic assumptions are true.

In the next part of the thesis we examine the role of entanglement in noisy quantum computations. Entanglement is necessary for noiseless quantum computers to have any quantum advantage, but it is not known whether the same is true for mixed state quantum computers. We show that entanglement, unexpectedly, does play a crucial role in the most well known mixed state computer: the one clean qubit model.

Finally, we investigate how closely classical simulation is related to another idea of classicality. This notion captures how easily the final state of a computation can be learnt, given samples of measurements from it. We find an extra condition under which a circuit that is classically simulable is also efficiently learnable.

Contents

Chapter 1

Classical simulation

1.1 Introduction

Once upon a time a computer was not a machine, it was a human who did computations; someone who solved algorithmic mathematical problems that varied from interpreting astronomy data to calculating rocket trajectories. When Turing borrowed the term 'computer' for his machine he did so intentionally. The Turing Machine was designed to mimic a human computer and automate this sort of mathematical problem solving.

Mechanical computers though don't just mimic some types of intelligent thinking. The human brain itself appears to be a complex computer. That's because the concept of 'computation' doesn't depend on hardware. Any system that can perform and certain logical steps is a computer. Because every computer is based on the same principles of logic, it was hypothesised that all computers are roughly equivalent. An algorithm can be translated to run on any model of computation and have roughly the same number of steps in every hardware. This is the Church-Turing hypothesis. It is not provable but was widely believed to be true.

Quantum computers are yet another hardware for computation, a hardware that utilises quantum mechanics. Quantum computers, though, are not just a mechanical brain. The ways of problem solving they can use are not accessible to either humans or classical computers. In some cases quantum computers arrived at an answer via fundamentally different methods, but they also seem to do it faster than any method possible classically [35, 79, 24, 3, 26]. If quantum computers really do have a speedup over classical computers then this is clear violation of the Church Turing hypothesis, and it would imply that quantum computers perform some logic that is fundamentally not equivalent to what we, as classical computers, use.

What is so special about quantum mechanics to allow this? What parts of the theory are responsible for this quantum advantage?

Finding the source of the quantum computing advantage is a large open problem in quantum computing. While in some sense we exactly what a quantum computer is and how they work, we don't have a high level understanding of what makes them work. What features of quantum mechanics are vital to what quantum computers do?

One way of attacking the problem is to consider the question *"when is a quantum computer efficiently classically simulable?"* Informally, a class of quantum computations is efficiently classically simulable when the computational problems it solves are also tractable on a classical computer. A quantum computation that has a quantum advantage, on the other hand, is one that cannot be efficiently simulated. In this thesis we will study computers which transition from having a quantum advantage to not having one by removing *one* resource. This will allow us to say that this property is resourceful and identifies a property of quantum mechanics that is at least necessary for a quantum speedup, relative to the background resources.

Identifying that a resource is necessary for a quantum speedup is not the same as proving that that resource is what is being utilised by a quantum computer for a speedup. It is a much more tractable and mathematically precise problem though. Answering it would be valuable to us in two ways. The first is practical. At the moment we still have very limited tools for diagnosing whether a quantum algorithm in fact does not have a speedup. Finding necessary features for a speedup will help us check whether this is the case. For the second reason, we draw an analogy with Bell's theorem. Bell's theorem showed that there was a particular feature of quantum mechanics that could be mimicked by any local theory of physics. It also identified that entanglement was necessary for quantum mechanics to show this behaviour. This alerted us to the idea that entanglement is an especially important feature of quantum theory. We hope that by identifying resources that are necessary for quantum computers we will also identify these as important features of quantum mechanics itself. This could lend us a new light for understanding fundamental physics.

In this thesis we take some modest steps toward when efficient classical simulations of quantum computers exist, while bearing in mind the loftier goals for this line of research.

1.1.1 Contributions of this thesis

A class of quantum computations is efficiently classically simulable when the computational problems it solves are also tractable on a classical Turing machine. A quantum computation that has a quantum advantage, on the other hand, is one that cannot be efficiently simulated. In

this thesis we will study computers which transition from having a quantum advantage to not having one by removing one resource. This will allow us to say that this property is resourceful and identifies a property of quantum mechanics that is at least necessary for a quantum speedup, relative to the background resources.
In **Chapter 2** we prove an extension of the well known Gottesman–Knill theorem. The original theorem shows that stabiliser circuits with stabiliser inputs are classically simulable. The extension considers the case of stabiliser circuits which have access to so-called magic state resource states. We show that in a sense it is possible to disentangle the non-classical resource (the magic states) from the classical part (the stabiliser circuit). The quantum computation can be simulated by a smaller quantum computer by using some classical polynomial time processing.

We then use this extension in **Chapter 3** to prove a class of circuits, Clifford Magic, cannot be classically simulated under plausible conjectures. Results of this type were first shown in References [23, 3, 27], for the classes IQP and BosonSampling. We show that Clifford Magic circuits can emulate certain classes, including IQP, and hence inherit their hardness properties.

In **Chapter 4**, we study another resource: entanglement. It is known that removing entanglement from pure state quantum computation makes the computation classically simulable [51, 88]. However the same result is not known for mixed state quantum computations: computations that allow mixed state inputs. The one clean qubit model is a quantum computer which takes the maximally mixed state and one pure qubit as input, and it has been shown that there is a limited amount of entanglement in this model; there does not need to been entanglement between the mixed and pure register [74], and (by some measures) the amount of entanglement in the model is bounded by a constant independent of the number of qubits [33]. We show that, despite this, entanglement is crucial in this model because without it the one clean qubit model is efficiently classically simulable. This lends weight to the conjecture that entanglement is in fact necessary for mixed state quantum computation to have a quantum advantage.

In **Chapter 5** we studied the role of entanglement as a resource in a different context to classical simulation, called quantum state learning. In the task of quantum state learning one receives some data about measurements performed on a state and, using that, must make predictions on the outcomes of unseen measurements. While the amount of data required to learn the state only grows linearly in the number of qubits [2], using that data to compute a hypothesis state is generally computationally intractable, classically and quantumly. However, it has been shown that learning can be performed efficiently for states that are generated by Clifford circuits, using many of the tools of the Gottesman–Knill simulation algorithm. This naturally leads to the question, how does efficient state learnability compare with efficient classical simulation? We introduce an extra condition on top of classical simulablity that guarantees efficiently learnability.

To illustrate this we prove two new examples of efficient learnability: states with low (Schmidt rank) entanglement and states described by an 'efficient' ontological model. An ontological model is a hidden variable model for quantum computation. We showed that when we restrict to quantum computations that can be efficiently simulated via an ontological model description that these states also become learnable.

1.2 Efficient classical simulation

Here we will review definitions important to this thesis, including the definitions of classical simulation that we will use through out.
When we ask whether a class of quantum circuits is classically simulable, we are actually concerned with whether a uniform family of circuits in that class are simulable. Informally, a family of circuits is uniform if it is not difficult to compute what circuits in the family look like. This is important, because if a family of quantum circuits cannot be efficiently classically described then it certainly cannot be efficiently classically simulated.

Definition 1.2.1. *A uniform family of quantum circuits is a mapping $s \to C_s$ where s is a n-bit string and C_s is a classical description of a circuit on n qubits. The mapping $s \to C_s$ is computable in time polynomial in n on a classical computer.*

In practice C_s will often refer to the description of the circuit as well as the circuit itself. We define a quantum circuit as follows.

Definition 1.2.2. *An* adaptive quantum circuit *C on n qubits, with input state α and output distribution P_C comprises the following ingredients. We have a specified sequence of steps (on the n-qubit state α) of length poly(n), with the following properties:*
(i) each step is either a unitary gate or a non-destructive Z basis measurement. Post-measurement states from intermediate measurements may be used as inputs to the next step.
(ii) each step is specified as a function of previous measurement outcomes by a classical (possibly randomised) poly(n) time classical computation.
If no steps depend on previous measurement outcomes then the circuit is called non-adaptive*, and if there are no intermediate measurements steps, then the circuit is called* unitary.

The output distribution P_C is the probability distribution of a specified set of measurements (called output measurements). Without loss of generality this may be taken to be the set of all measurements of the circuit C and we often omit explicit mention of the output set.

The definition of efficient simulation we will use most often in this work is the notion of weak simulation. When considering whether a computer can simulate a family of quantum circuits, it

is most natural to consider whether or not the classical computer can perform a weak simulation of the quantum circuits. If it can, it can produce outputs that are drawn from the same distribution as the original, and in approximately the same time. Therefore, if one of these two computers where put in a black box, it would be impossible to guess which it is by only looking at the outputs.

Definition 1.2.3. *We say that a circuit C (on n qubits, with input state α, and output distribution P_C) from a uniform family can be* weakly simulated *by a circuit $\tilde{C}$ (on m qubits, with input state β, and output distribution $P_{\tilde{C}}$) if*
(i) a description of the circuit $\tilde{C}$ may be given by a classical poly(n) time (possibly randomised) translation from a description of C, and
(ii) a sample of the distribution P_C can be produced from a sample of $P_{\tilde{C}}$ together with poly(n) time classical (randomised) computation.

However, even this definition of classical simulation is slightly too strong; a physical quantum computer with any amount of noise would not be able to weakly simulate an ideal quantum computer. That is because we have not allowed for any error tolerance in the above definition. The definition of an ε-weak simulation allows for error in the simulation, and is therefore the physicially relevant definition of a simulation.

Definition 1.2.4. *We say that a circuit C (on n qubits, with input state α, and output distribution P_C) from a uniform family can be* weakly simulated *by a circuit $\tilde{C}$ (on m qubits, with input state β, and output distribution $P_{\tilde{C}}$) if*
(i) a description of the circuit $\tilde{C}$ may be given by a classical poly(n) time (possibly randomised) translation from a description of C, and
(ii) a sample of the distribution P'_C can be produced from a sample of $P_{\tilde{C}}$ together with poly(n) time classical (randomised) computation, where $||P'_C - P_C|| \leq \varepsilon$

In this work we will often say a computer is classically simulable or efficiently classically simulable to mean it can be (ε)-weakly simulated by a classical computer.
Sometimes though, we will consider *strong simulations*. In this notion, the simulator does not produce a sample from the distribution P_C. Instead, for any possible output bit string x, the simulator can produce the probability of x. This probability can be produced to k digits of accuracy in poly(n,k) time. Strong simulation do not necessarily imply weak simulation unless the probability can also be computed for the marginal distributions of P_C. See Proposition 1 of [82] for a proof of this fact.

1.3 Review of classical simulation results

In this section of the thesis we review some of the important results in classical simulation. We also contextualise the results presented in this thesis which will explain why these questions

where considered in the first place.

The Gottesman–Knill theorem was one of the first nontrival classical simulation results [41]. The restricted class of circuits it considers are those with stabiliser inputs, and adaptive Clifford gates (defined in Section 2.1), and the theorem shows that this restricted class is strongly classically simulable. One reason this result is particularly surprising is these stabiliser circuits are in fact useful for many other applications in quantum information and computation, for example quantum teleportation and quantum error correction [42]. Furthermore, adding almost any other 1 qubit gate to the Clifford gates makes the set universal [65], so it is surprising that the slight restriction to Clifford gates makes the computer simulable.

Though there are many 1 qubit gates we could choose to add to make stabiliser circuits universal, the canonical choice is the T gate (defined in Section 2.1). That is because a T gate can be implemented using a 1-qubit "magic state" $|A\rangle$ and an adaptive stabiliser circuit– a construction called a T-gadget [16]. This means we can think of the state $|A\rangle$ as the resource that lifts stabiliser circuits from classical to universal quantum computation.

We want to find the minimum resources necessary for stabilisers to become truly quantum. Identifying this would show us exactly what exactly stabilisers are missing to be truly quantum. That is why it is interesting to ask about various restrictions of stabiliser circuits. What happens, for example, when we allow access to magic states (or more generally any product input state), but restrict adaptivity? This means we cannot perform intermediate measurements nor use them to inform which gates to apply next. Therefore, even with access to magic states, we cannot perform T-gadgets. Ref [50] considers questions of this type, with the results summarised in Figure 1. Ref [56] considers further extensions of this work using different notions of classical simulation, and Ref [29] considers when the inputs are mixed states.

The figure shows that a stabiliser circuit with general product state input but no adaptions cannot be efficiently weakly simulated. One thing to note however is that that theorem is for weak simulation *without error* as per Definition 1.2.3. As we noted in the above section, it is much more natural to allow the weak simulation to have error, as in Definition 1.2.4. We prove that the result still holds with error in **Theorem 3.4.3**.

This means that stabiliser circuits with magic state inputs but no adaption is truly quantum; it cannot be classically simulated. This is despite the fact that this class does not seem to be universal- without adaption it does not seem possible to use magic states for performing non Clifford gates. This model is therefore likely to be somewhere between universal quantum and classical computation. This raises an important point; there is likely no ‘gap’ between universal

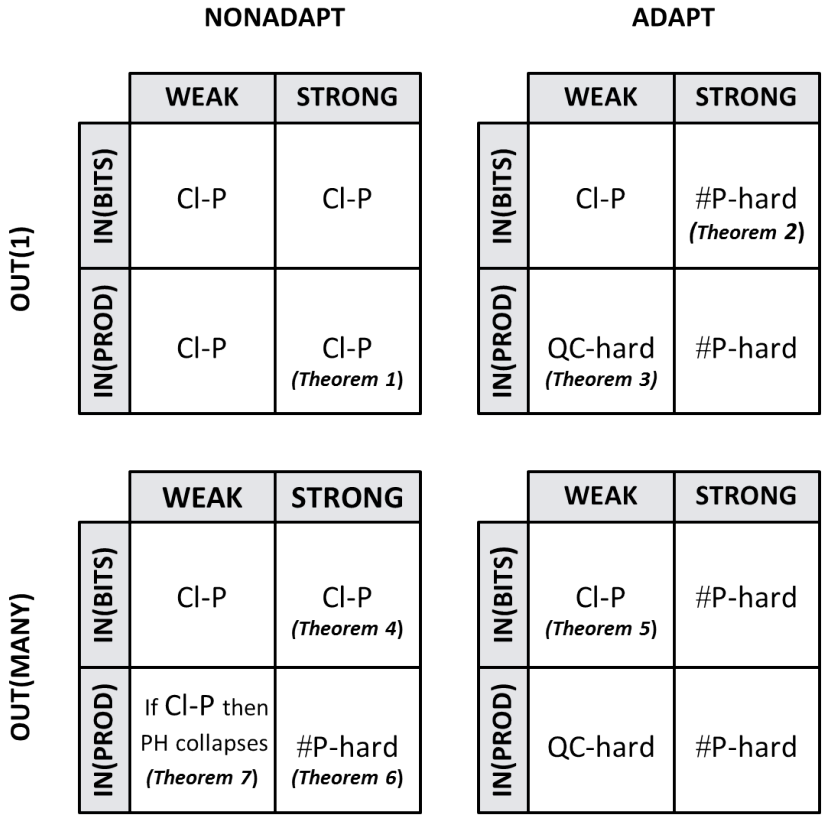


Figure 1.1 This figure is from Ref [50]. The possible restrictions of stabiliser circuits considered are 1) whether or not magic inputs are allowed (labelled in(prod) and in(bits) respectively), 2) whether one or many lines are measured at the end of the computation (labelled out(1) and out(many) respectively), 3) whether adaption is allowed or not (labelled adapt and nonadapt respectively) and 4) whether the classical simulation is required to be strong or weak. If a particular combination is marked with "Cl-P", this means the required simulation takes polynomial time on a classical computer. "QC-hard", in contrast, means this combination is actually as hard to simulate as a general quantum computer. "#P-hard" means that performing the simulation is at least as hard any problems in the class #P (see [69] for a formal definition of this class). "If Cl-P then PH collapses" means that if this can be simulated in polynomial time, then a widely conjectured complexity theorem would be false (see Section Chapter 3 for more on this). Therefore this is unlikely to be true. The theorems referenced in the table are from Ref [50].

quantum and classical computation, but instead a spectrum of intermediates.

Access to polynomially many magic states makes stabiliser computations go from classically simulable to universally quantum. But does having an intermediate amount of magic states result in an intermediately powerful computation? References [4, 18, 17] showed that if there are t magic states in a stabiliser computation then it can be simulated in time exponential in t.

Therefore if the number of magic states is growing only logarithmically, the computer is still classically simulable. It is only when one has more magic states than this makes the simulation difficult. But does adding more magic states make a quantum computer progressively stronger? To understand this we need to look beyond purely classical simulation. In Ref [22] it is shown that any stabiliser circuit with t magic states can be simulated by another stabiliser circuit with only t magic state inputs (and no other qubits) with the help of polynomial time classical computation. This suggests that the strength of the computations depends only on the number of magic states and not the total number of qubits available. However, in that result the new quantum computation is not a standard one using gates. Instead the computation is performed by t qubit entangling Pauli measurements, which are not generally considered a standard resource. In **Theorem 2.2.1** we show that this measurements can be replaced by standard 2 qubit Clifford gates instead. Therefore we can in fact conclude that the power of the computation depends on the number of magic state inputs, not the total number of qubits.

Magic states then seem to be an important resource in quantum computing, relative to stabiliser circuits. However, the state $|A\rangle$ is not the only type of state with this magic. There are other states that can be used to make non-Clifford gate gadgets. The magic states evidently cannot be stabiliser states because a stabiliser circuit with those inputs is classically simulable. Which other input states make the circuit simulable, and which do not? For those that are truly "magic", what is the source of the magic?

One suggestion was that magic states are those that display contextuality. Contextuality is a property of hidden variable models of quantum mechanics [60]. A hidden variable model is one in which the quantum state is not a complete description of the actual state of the particles. There is some true (ontic) state which we do not have access to, and the quantum state merely represents the ensemble of ontic states over which we have classical uncertainty. Kochen–Specker shows that hidden variable models must have contextuality; the outcome of a measurement on one these underlying ontic states must depend on what other compatible measurements are performed simultaneously. For example, suppose there are two spin half particles described by a hidden variable model and they are known to be in a particular ontic state. Then to predict whether the spin of the first particle will up in the Z direction when measured, we need to know whether the second particle's Z spin will also be measured, or if instead Z_1Z_2 will be measured etc.

For stabiliser states of odd dimension it is possible to construct a non-contextual hidden variable model, and using that method there is an alternative way to simulate these states [**?**]. This suggests that simulation and non-contextual hidden variables may be related. Hidden variable models that are not constrained to be contextual are not necessarily simulable. In fact it is

possible for a hidden variable model to reproduce quantum mechanics [12]. We wished to show that there is another constraint on hidden variable models (unrelated to contextuality) that makes them simulable. In **Theorem 5.2.7** we show that if the number of ontic states is polynomial, and certain functions are polynomial time computable, the hidden variable model is efficiently simulable.

Ref [48] showed that for stabiliser quantum computation on qudits of odd prime dimension contextuality is necessary for a quantum advantage; if the input states do not have contextuality (relative to stabiliser measurements), then the entire circuit can be classically simulated. This result cannot be shown for qubits because all qubit states are contextual relative to stabiliser measurements: a result shown using the well-known Mermin–Peres magic square [59, 73]. However, in the Ref [11, 45, 38] the authors consider the qubit case in which they restrict the measurements allowed in a natural way. Restricting the measurements this way still allows stabiliser circuits to be implemented as measurements (which can be seen using **Theorem 2.2.1**), but not all states are contextual under these measurements. In this case, they showed that contextuality is still necessary for a quantum advantage. Furthermore, it was shown that a property called negativity is essentially equivalent to contextuality [80], and negativity is known to be necessary for quantum computation [87, 86, 72, 71, 77].

It is natural to ask then if contextuality is both necessary and sufficient for quantum computation. Is it possible to use a polynomial number of any contextual state in a stabiliser circuit and get a quantum speedup? Recently it was showed that this is not the case. Contextuality (in the qudit case) is not sufficient. There are contextual states that nevertheless make any stabiliser circuit simulable.

Contextuality is not the only candidate for the source of the quantum computing advantage. Entanglement has also been implicated in that role, though like contextuality the evidence for it is mixed. Entanglement is necessary for pure state computation. Without it a quantum computer becomes simulable [51]. If the entanglement is very restricted (by some measures) then the computation is also simulable [51, 88, 58, 49]. In **Theorem 5.2.4** we consider states with this type of restricted entanglement, and show that they are classical in another sense; they can be efficiently PAC learned (see Chapter 5 for more on this).

However, Ref [84] pointed out that the measures considered in these works are not physical because they change discontinuously when a state changes smoothly. If entanglement is truly important in quantum computing, one would expect that restricting it (with some physically relevant measure) would decrease the effectiveness of the quantum computer. This has not been

shown.

Another way to probe the utility of entanglement would be to see how useful such a state is as an initial resource. This can be motivated in an analogous way to studying magic state inputs in stabiliser circuits. Stabiliser circuits cannot produce nonstabiliser states, so any 'magic' has to be supplied by the input. Similarly, we can consider a circuit that does not produce entanglement, so any entanglement must be supplied by the input state. This is the case for measurement based quantum computing, where the input is some n qubit entangled state (see Section 3.5 for more on measurement based quantum computing). The measurement based model only allows 1-qubit measurements, which is why it does not produce entanglement. The canonical input state in a measurement based computation is highly entangled, so one may hypothesis that most other entangled states ought to be useful input states as well. This is examined in Ref [43] where it is shown that computations using states with a large amount of entanglement are in fact not useful; any NP problem such a computer can solve in polynomial time can also be solved in polynomial time by a classical computer.

Entanglements role in quantum computing becomes even less clear when one considers mixed state quantum computing. Mixed state computation is where the input is a mixed, rather than pure, quantum state. Generally, we expect the power of mixed state quantum computers to be somewhere between pure state quantum computation and classical. Intuitively, adding more noise ought to decrease that that power. There are very few well studied mixed state models. The most well known is The One Clean Qubit Model. This model is striking because of how much noise it has; only one qubit is assumed to be pure, while the rest of qubits are maximally mixed. Despite having this much noise there is evidence that these computers cannot be classically simulated [64, 62, 39]. Entanglement is necessary for pure state quantum computation, but this is not known to be true for mixed state computers. It is possible that some mixed state computers have a quantum advantage without any entanglement at all. If this is the case, then that would conclusively show that entanglement is not the source of quantum speedups. The One Clean Qubit Model looked like a good candidate for finding such a counterexample. It is possible to arrange these computations so that there is no entanglement between the clean qubits and the noisy ones at any point in the computation. It is also known that there is a low amount of entanglement across any bipartition in the negativity of entanglement measure [33]. Therefore it seemed plausible that a One Clean Qubit computer without entanglement might still have some quantum advantage. In **Theorem 4.2.5** we show that this is in fact not the case. Even the One Clean Qubit Model needs entanglement. Perhaps then entanglement is at least always necessary for a quantum advantage.

Another intriguing topic in classical simulation is matchgate computation. Matchgates are a restricted class of gates and matchgate circuits are ones in which these gates are only applied between nearest neighbour qubits. These circuits are in fact classically simulable [83, 52, 28]. The states that can be created in a standard matchgate circuit from computation basis states are called Gaussian, and they can be efficiently described classically, and that description can be updated efficiently as the state evolves through a matchgate circuit. However, adding swap gates to a matchgate circuit (i.e. the ability to perform matchgates between distant qubits) makes the computer universal. Recently, magic states for matchgate computations were shown to exist [47]. In fact, all pure states that are non-Gaussian are magic states; adding any type of them makes a matchgate circuit universal. Unlike the stabiliser or measurement based computing cases above, it is known what types of inputs are necessary and sufficient.

Matchgate computations with computational basis inputs where shown to be equivalent to computation using noninteracting fermions for computation [54, 81, 36]. It is surprising that this case is classically simulable, because the same type of computer using bosons instead of fermions is known to be hard to classically simulate, under plausible assumptions [3]. Results of this type are referred to as quantum supremacy theorems (see Section 3.1) and the methods used to prove them are important tools for classifying intermediate quantum computers. There are other methods but they do not show exponential separations [19, 20, 89] or are oracle separations. There is yet no proof that quantum computers are more efficient than classical. Supremacy arguments show that this must be the case though, unless certain conjectures are false. They can also be used to show certain restrictions of quantum computers are also not simulable (for examples see Refs [27, 62, 15, 70, 61, 57, 44, 68, 9] and **Theorem 3.4.3**). This allows us to distinguish the truly quantum from classical computations.

Chapter 2

Gottesman–Knill Theorem

The Gottesman–Knill theorem showed that 'stabiliser quantum circuits' can be classically simulated [41]. This result is surprising because stabiliser circuits are capable of generating highly entangled states, including Bell states. And yet, despite being very 'quantum' in nature, they are still simulable. This shows that even if entanglement is an important resource in quantum computing it is not enough for it to be present in large amounts, it must be utilised correctly.

A stabiliser circuit is composed of Clifford gates, which is not a universal gate set. However, the addition of any one qubit gate outside of the set does promote the gate set to universality. A common choice is the 'T gate', defined below. Because Clifford circuits, with computational basis inputs, are classically simulable but Clifford+T circuits are universal, T gates appear to a resource for quantum computing. However, to strengthen such a claim, one would ideally want to show that the simulation complexity grows exponentially in the number of T gates. This would show that with too few T gates the circuit remains classically simulable. This result was shown in Ref [4] and improved in Ref [18].

In this chapter, we prove another generalisation of the Gottesman–Knill theorem, that allows one to, in a sense, separate the 'hard' part of a circuit that corresponds to the T gates from the 'easy' stabiliser part. The easy part is then simulated via efficient classical processing, and the remaining hard part that must be done quantumly is now smaller than the original circuit. This result, the Extended Gottesman–Knill Theorem, is stated in Theorem 2.2.1.

This theorem is conceptually interesting because of its hybrid quantum–classical nature. Simulation results are generally entirely classical and show that quantum computations with a limited amount of some quantum resource can be classically simulated efficiently. Some simulation algorithms, such as [4, 51, 88] also extend to simulating general quantum algorithms (inefficiently), and show that the simulation speed is limited by the amount of the quantum resource present. The algorithm provided by Brayvi, Smith and Smolin [22], in contrast, is neither

purely classical or quantum. They showed the classical and quantum parts of the circuit can be separated in the stabiliser case and run on a classical and quantum computer respectively. However, the quantum part of their algorithm needed to run on a quantum computer with n qubit entangling measurements. This is a standard resource in quantum computing. We showed that this is in fact not necessary. The quantum part of their algorithm can be performed on a standard quantum computer with Clifford gates and 1 qubit measurements.

This chapter is organised as follows. In Section 2.1 we review the stabiliser formalism and the original Gottesman–Knill theorem.

2.1 The stabiliser formalism

The n qubit Pauli group, $\mathcal{P}_n$, is multiplicatively generated by tensor products of the 1-qubit Pauli operations. In general, a n qubit Pauli operator $P \in \mathcal{P}_n$ is in the form $P = (i)^x P_1 \otimes ... \otimes P_n$, where $x \in \{0,1,2,3\}$ and P_j is a single qubit Pauli X, Y, Z or $\mathbb{I}$. This means an element of this group can be described with $2n+2$ bits of information.

A stabiliser group is a subgroup of $\mathcal{P}_n$ such that all elements commute with each other and such that $-\mathbb{I}$ is not in the group. Though these groups can be exponentially large they can be characterised by at most n independent generators. A Pauli P is said to be dependent on Paulis $Q_1,\ldots,Q_K$ if $P = \pm Q_1^{a_1} \ldots Q_K^{a_K}$ for some $a_1,\ldots,a_K \in \{0,1\}$. A stabiliser group is generated by at most n independent Pauli operators, which we will denote as $\mathcal{S} = \langle S_1,...,S_k \rangle$, where $k \leq n$.

A pure stabiliser state $|\psi_S\rangle$ is a state that is uniquely described by a stabiliser group $\mathcal{S}$ with n independent generators in the following way. $S|\psi_S\rangle = |\psi_S\rangle$, for all $S \in \mathcal{S}$. It is so called because it is stabilised by all elements of the group.

More generally an n qubit state ρ mixed stabiliser state is uniquely defined as the uniform mixture of all pure stabiliser states stabilised by $\mathcal{S} = \langle S_1,...,S_k \rangle$, where $k \leq n$. Alternatively, it is the state that is obtained after measuring the maximally mixed state with (commuting) Pauli measurements S_1,..., S_k and post selecting outcome $+1$:

$$\rho = \frac{1}{2^{n-s}} \prod \frac{\mathbb{I} + S_i}{2}. \tag{2.1}$$

Because a stabiliser group uniquely describes a stabiliser state, such a state can be described with less than $(2n+1)n$ bits of information[1].

Clifford unitaries are defined by their commutation relationship with Pauli operators. A unitary U is Clifford if, for any $P \in \mathcal{P}_n$

$$PU = UP', \tag{2.2}$$

where $P' \in \mathcal{P}_n$. Therefore commuting a Clifford unitary and a Pauli unitary is possible, at the cost that the Pauli will be transformed into another element of the group $\mathcal{P}_n$.
Unitary Clifford circuits will always be assumed to be given as circuits of some chosen set of one and two qubit Clifford gates that suffice for any Clifford operation e.g. the Hadamard gate H, controlled NOT gate CX and phase gate $S = \mathrm{diag}(1\ i)$.

We will also consider adaptive Clifford circuits with intermediate Z measurements and possibly adaptive choices of later gates, as formalised in Definition 1.2.2.

Theorem 2.1.1. *(Gottesman–Knill theorem [41, 4]) An adaptive Clifford circuit with stabiliser input can be classically simulated in the sense of Definition 1.2.3.*

This theorem is proven by showing that the state throughout this computation remains a stabiliser state and therefore it can be described efficiently; the stabilisers can be updated efficiently each step; and measurement probabilities can be computed efficiently. See section 10.5 of Ref [67] for details of the proof.

We will use the non-Clifford T gate defined by $T = \mathrm{diag}(1\ e^{i\pi/4})$. It is well known that the T gate can be implemented by the so-called T-gadget [66], using an extra ancilla qubit line (labelled a) in 'magic' state $|A\rangle \frac{1}{\sqrt{2}}(|0\rangle + e^{i\pi/4}|1\rangle)$ and adaptive Clifford operations: to apply T to a qubit line k in a circuit, we first apply CX_{ka} with the ancilla as target qubit, and then measure the ancilla qubit in the Z basis giving outcome $+1$ or -1 (always with equal probability). Finally an S correction is applied to the original qubit line if the outcome was -1. The ancilla qubit is never used again and may be discarded. The final result in every case is to apply T to line k up to overall phase. It will also be useful to note that we can implement the $T^\dagger$ gate using a similar gadget: we perform the T-gadget process as above but for the final adaptive correction we instead apply an S^3 correction if the outcome was $+1$.

Clifford operations with T gates are universal for quantum computation. Using the T-gadget we see that any (general universal) circuit composed of Clifford gates and a number t of T gates can be rewritten as an adaptive circuit of only Clifford gates (and intermediate Z basis measurements) with the addition of t additional ancilla qubit lines initialised in state $|A\rangle^{\otimes t}$.

[1]The elements of a stabiliser group cannot have i or $-i$ as a phase. Otherwise such an element multiplied by itself gives $-\mathbb{I}$, which is not allowed to be in the group.

2.2 Extending the Gottesman–Knill Theorem

We begin by establishing an extended form of the Gottesman–Knill theorem that will be used later in our development of CM circuits.

As noted above, universal quantum computation can be performed using adaptive Clifford circuits which include additional (non-stabiliser) $|A\rangle$ state ancilla inputs, motivating the consideration of Clifford circuits on such more general inputs. In our extension of the Gottesman-Knill theorem we consider adaptive Clifford circuits but now allow the input to have a non-stabiliser part. We show that it may be weakly simulated by a hybrid classical-quantum process whose quantum part (obtained by an efficient classical reduction from the description of the original circuit) is an adaptive Clifford circuit acting now only on the non-stabiliser part of the original input, thereby relegating the stabiliser-input part of the original computation into efficient classical computation instead. In the special case where the initial input is fully a stabiliser state, we recover the standard Gottesman–Knill theorem, as our hybrid process then has no residual quantum part. This is stated formally as follows:

Theorem 2.2.1. *(Extended Gottesman–Knill Theorem) Let $\mathcal{C}$ be any adaptive Clifford circuit with input state $\sigma \otimes \rho$, where σ is a stabiliser state of n qubits and ρ is an arbitrary state of t qubits, and the output is given by measurement of any specified qubit lines. (Usually we will also have $t = O(\mathrm{poly}(n))$). Then*
(i) $\mathcal{C}$ can be weakly simulated by an adaptive Clifford circuit $\mathcal{C}^*$ on t qubits with input ρ, assisted by poly$(n+t)$-time classical computation, and with $\mathcal{C}^*$ having at most t (intermediate or final) measurements;
(ii) if $\mathcal{C}$ is non-adaptive then $\mathcal{C}^*$ may be taken to be unitary (with Z basis measurements only for outputs at the end).
(iii) If some Z measurements in C are to be postselected to outcome $+1$, this circuit can be weakly simulated by a circuit $\mathcal{C}^*$ as in case (i), where some of the Z measurements are postselected to outcome $+1$.

The proof of the Extended Gottesman–Knill Theorem will be given in Subsection 2.3 below. It rests on the so-called Pauli based model of computation (PBC) introduced by Bravyi, Smith, and Smolin in [22]. Before the proof of Theorem 2.2.1 we will in Subsection 2.2.1, give an account of (a slightly generalised version of) the PBC formalism and its main features that we will use.

The Extended Gottesman-Knill theorem will be used in this paper to show that certain quantum circuits can be simulated by CM circuits (cf Section 3.3). However, we expect that the theorem will be of independent interest, for example for considerations of compiling quantum circuits with as few qubits as possible. Indeed starting with the circuit model of quantum computation we may represent any circuit as a circuit of Clifford gates and T gates, and then use T-gadgets

to implement the T gates, resulting in an adaptive Clifford circuit. Implementing the circuit this way allows for error correction using stabiliser codes [66], but it also increases the number of qubits. Given the high practical cost of adding extra qubits, one naturally strives to minimise their number in near term devices. The Extended Gottesman–Knill theorem provides a way to remove all qubits originally in a stabiliser state, as well as any stabiliser ancillas. The resulting circuit is also an adaptive Clifford circuit, now having at most t measurements. This is summarised in Figure 1.

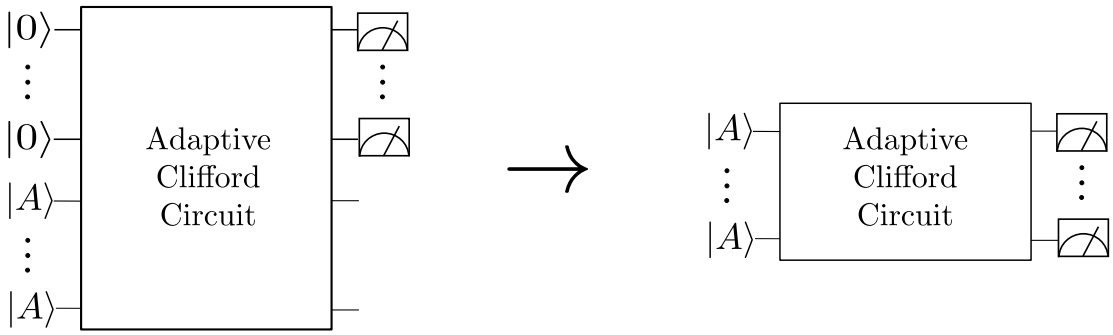


Figure 2.1 The Extended Gottesman Knill theorem (Theorem 2.2.1) allows us to take a universal quantum circuit expressed as a Clifford circuit with T-gadgets and compress it using only a classical polynomial time overhead. This compression removes all input state components that are stabilisers and the resulting circuit is an adaptive Clifford circuit with a number of (intermediate and final) measurements at most equal to the number of lines in the compressed circuit.

In [5] and [18] a different kind of extension of the Gottesman-Knill theorem is developed. It is shown that a circuit on n qubit lines with stabiliser input and t T gates, can be classically simulated in time exponential in t and polynomial in n. This reduces to the original Gottesman–Knill theorem when $t = 0$. Our Extended Gottesman Knill theorem provides an alternative proof of this fact: using Theorem 2.2.1 any such computation (after replacing T gates by T-gadgets) can be compressed to a quantum computation on t qubits, and this can be and then be classically simulated in time exponential in t.

2.2.1 The Pauli based model of computation (PBC)

Definition 2.2.2. *(PBC circuits and the Pauli based computing model)*
(i) A PBC circuit *C on t qubits with any input state ρ, is a sequence C of pairwise commuting and independent Pauli measurements $P_1, \ldots, P_s$ from $\mathcal{P}_t$ (applied sequentially to ρ with each post-measurement state being available for the next measurement). The choice of each P_i can generally adaptively depend on previous measurement outcomes. If no P_i depends on previous measurement outcomes then the PBC circuit is called non-adaptive.*

*(ii) For computational applications (*the PBC model of computing*) we will use a uniform family* $\{C_w : w \in \mathcal{B}\}$ *of PBC circuits on* $t = \text{poly}(n)$ *qubits where n is the length of the bit string w, and furthermore, each* C_w *is required to have the input state* $\rho = |A\rangle^{\otimes t}$*. The result of the computation is given by a specified poly*(n) *time (randomised) classical computation on w together with the measurement outcomes of the circuit* C_w*.*

Theorem 2.2.3. *(adapted from Ref [22]). Let C be any (generally adaptive) quantum circuit on* $n+t$ *qubits with input state* $\alpha = \sigma \otimes \rho$ *where* σ *is a stabiliser state of n qubits and* ρ *is any state of t qubits. Suppose also that the unitary steps of C are all Clifford gates. Then:*
(i) C may be weakly simulated by a (generally adaptive) PBC circuit $\tilde{P}_1, \ldots, \tilde{P}_s$ *on t qubits with input state* ρ*, and with* $s \leq t$ *steps.*
(ii) If C is non-adaptive (with final Z basis measurement outputs) then the PBC circuit $\tilde{P}_1, \ldots, \tilde{P}_s$ *in (i) can also be chosen to be non-adaptive.*
(iii) If some Z measurements in C are to be postselected to outcome $+1$*, then this circuit can be weakly simulated by a PBC circuit in which some of the Pauli measurements are postselected to outcome* $+1$*.*

We give the proof in full (following the method of [22] and extending the latter for clauses (ii) and (iii) above) dividing it into labelled sections. We begin with two supporting lemmas.

Lemma 2.2.4. *[22] Let* $P, Q \in \mathcal{P}_n$ *be anti-commuting Pauli operations and let* $|\psi\rangle$ *be an eigenstate of P with* $P|\psi\rangle = \lambda_P|\psi\rangle$*,* $\lambda_P = \pm 1$*. Then:*
(i) Measurement of Q on $|\psi\rangle$ *gives result* $\lambda_Q = \pm 1$ *with equal probabilities half.*
(ii) The operator $V(\lambda_P, \lambda_Q) = (\lambda_P P + \lambda_Q Q)/\sqrt{2}$ *is always a unitary Clifford operation.*
(iii) $V(\lambda_P, \lambda_Q)|\psi\rangle$ *is the normalised projection of* $|\psi\rangle$ *onto the* λ_Q*-eigenspace of Q.*
Hence measurement of Q on $|\psi\rangle$ is equivalent to classically choosing (offline) a uniformly random $\lambda \in \{-1, +1\}$ and applying the Clifford unitary $V(\lambda_P, \lambda)$ to $|\psi\rangle$.

Proof. We have $|\psi\rangle = \lambda_P P|\psi\rangle$.
For (i) we have Prob (Q measurement gives ± 1) $= \left|\left|\frac{1}{2}(I \pm Q)|\psi\rangle\right|\right|^2$. Replacing $|\psi\rangle$ by $\lambda_P P|\psi\rangle$, and using the fact that $PQ = -QP$ and that P is unitary, we readily see that the two probabilities are equal.
For (ii), using $P^2 = Q^2 = I$ and $PQ = -QP$ we can check directly that $V(\lambda_P, \lambda_Q)V(\lambda_P, \lambda_Q)^\dagger = I$. Similarly for any Pauli R, for each of the four possible combinations of R commuting or anti-commuting with P and Q, we can check directly that $V(\lambda_P, \lambda_Q) R V(\lambda_P, \lambda_Q)^\dagger$ is a Pauli operation (being just a suitable product of P, Q and R in each case).
For (iii) the normalised post-measurement state after outcome λ is

$$\frac{(I + \lambda Q)}{\sqrt{2}}|\psi\rangle = \frac{(\lambda_P P + \lambda Q)}{\sqrt{2}}|\psi\rangle = V(\lambda_P, \lambda)|\psi\rangle .$$

□

We will also use the following fact which is easily checked.

Lemma 2.2.5. *For any $P = \pm A_1 \otimes \ldots \otimes A_n \otimes B_1 \otimes \ldots \otimes B_t \in \mathcal{P}_{n+t}$ with all A_i's and B_j's being X, Y, Z or I, write $\tilde{P} = \pm B_1 \otimes \ldots \otimes B_t \in \mathcal{P}_t$ (with same overall sign as P). If P commutes with $Z_1, \ldots, Z_n \in \mathcal{P}_{n+t}$ then each A_i is either Z or I. If for all i, each A_i is either I or Z, then for any t-qubit state $|\psi\rangle$, the measurement of P on $|0\rangle^{\otimes n}|\psi\rangle$, and the measurement of $\tilde{P}$ on $|\psi\rangle$, give the same output distributions and corresponding post-measurement states of the form $|0\rangle^{\otimes n}|\psi'\rangle$ and $|\psi'\rangle$ respectively, with the same t-qubit states $|\psi'\rangle$.*

Proof of Theorem 2.2.3
Let $\mathcal{C}$ be any adaptive circuit whose steps are either unitary Clifford gates or Z measurements, with K measurements in total. For clarity, we will give the proof for the case where σ is the pure state $|0\rangle^{\otimes n}$. The general case of arbitrary (mixed) stabiliser state σ is proved similarly by just replacing $Z_1, \ldots, Z_n$ in (b) below by a set of generators $S_1, \ldots S_r$ $(r \leq n)$ of the stabiliser group defining σ.

(a) Starting with the rightmost Clifford gate and working successively to the left, we commute each gate out to the end of the circuit beyond the last measurement. As a result each Z measurement will become conjugated into a Pauli measurement $P_i \in \mathcal{P}_{n+t}$ which may be efficiently determined. Unitary gates applied after the measurements have no effect on the outcomes so we delete them, and we are left with a sequence $P_1, P_2, \ldots, P_K$ of (generally adaptive) Pauli measurements (where s is the number of Z measurements in $\mathcal{C}$), acting on input state $|0\rangle^{\otimes n} \otimes \rho$. Remark on (a): we could instead commute out the Clifford gates in sections, interleaved with the process to be described in (c) below, as follows. As we consider each successive measurement Q_i of the original circuit in turn (working from the leftmost one) we commute only the Clifford gates on the left of Q_i to the right of it, and staying to the left of the next measurement, to obtain P_i as above, and then apply (c) to P_i. All gates are thus eventually commuted out beyond the last measurement as we consider each measurement in turn. This commuting process interleaved with (c) has the advantage that for adaptive gates (depending on previous measurement outcomes) the identity of the gate is always fixed before it is commuted to the right, and we never need to carry forward any variables of adaptation.

(b) Next we prefix the sequence in (a) with "dummy" Z measurements for each of the first n lines obtaining the list

$$(\text{LIST}): \quad Z_1, Z_2, \ldots, Z_n, P_1, P_2, \ldots, P_K.$$

This has no effect as the input is $|0\rangle$ on each of these lines (and the Z measurements all give result $+1$ with certainty).

(c) We now define our PBC process. We have a t-qubit register initially in state ρ. Looking at (LIST) in (b) we work successively through the P_j's starting with P_1(not the dummy Z's). For each P_j:
(i) If P_j is dependent on measurements already performed (which may be efficiently determined [66]), delete P_j from (LIST) and just calculate its outcome from previous recorded measurement results. Move to the next measurement in (LIST).
(ii) If P_j commutes with all measurements to the left in (LIST) (including the dummy Z's too), measure $\tilde{P}_j$ (as in Lemma 2.2.5) on the register and record its value λ_{P_j}. Then move to the next measurement in (LIST).
(iii) If P_j anticommutes with some measurement N (possibly a dummy Z) on the left (which had outcome λ_N), classically randomly choose $\lambda_{P_j} \in \{+1,-1\}$ and record it. Then delete P_j from (LIST) and replace it by the unitary Clifford $V(\lambda_N, \lambda_{P_j})$ (as in Lemma 2.2.4). Then update (LIST) by commuting out $V(\lambda_N, \lambda_{P_j})$ to the right. By Lemma 2.2.4 this process simulates the P_j measurement and its post-measurement state for subsequent measurements. Then move to the next measurement in (LIST).

It is clear that when we have treated all P_j's in (LIST) we will have performed a list of $s \leq K$ measurements on the t-qubit register, which are independent and commuting Pauli measurements (the only quantum action on the register occurring in (ii)), and this process is assisted by efficient randomised classical computation. Since the measurements are all independent and commuting, we must have $s \leq t$.

Independently of actually implementing the measurements on the quantum register, the process described in (c) above provides an efficient classical (generally randomised) procedure which, given a sequence of measurement outcomes $m_1, \ldots, m_l$ up to any stage l, determines the next quantum measurement that's guaranteed to be independent of all previous measurements and commuting with them i.e. a bonafide PBC circuit. This completes the proof of Theorem 2.2.3(i).

(d) We now prove Theorem 2.2.3(ii). If $\mathcal{C}$ is non-adaptive then we may assume without loss of generality that it is a unitary circuit U followed by final measurements $Z_{i_1}, \ldots, Z_{i_s}$ on specified qubit lines $i_1, \ldots, i_s$ [53]. Then in (b) we will obtain the non-adaptive list $Z_1, Z_2, \ldots, Z_n, P_1, P_2, \ldots, P_s$. Here $P_k = UZ_{i_k}U^\dagger$ for $k = 1, \ldots, s$, which are commuting and independent. However some may anticommute with an initial dummy Z measurement. Then following the process of (c)(iii) (with P_j and N as in (c) above), N must be one of the dummy Z's, whose measurement outcome $\lambda_N = +1$ is deterministic. Thus the unitary gate $V(\lambda_{P_j}, \lambda_N)$ involves no adaptations, and the sequence remains non-adaptive after $V(\lambda_{P_j}, \lambda_N)$ is commuted out to the end (although it depends on the classical random choice of λ_{P_j} that can have been chosen a priori). Continuing in this way, we note that if any subsequent updated operator M anticommutes with any earlier operator N,

then M must always anticommute with one of the dummy Z's too. This is because at any iteration stage, the operators after the dummy Z's are given by initial P_i's conjugated some number of times by operators V that are always in the algebra generated by the P_k's and dummy Z's (i.e. the successive V's that have been commuted out). Thus if M commuted with all the dummy Z's, it must also commute with all preceding operators N (recalling that the P_k's were all commuting).

Now by choosing an anticommuting N to always be a dummy Z, λ_N will always be $+1$ and no adaptation is ever introduced by (c)(iii) so, since the initial list of P_i's was non-adaptive, the final PBC process will be non-adaptive too. This proves Theorem 2.2.3(ii).

(e) Finally we prove Theorem 2.2.3(iii). In the case of postselection we proceed with all the steps as above as though there was no postselection, except (c)(iii). Suppose that the measurement P_j in that step is postselected to outcome $+1$. In that case, do not randomly choose λ_{P_j}, but set it to $\lambda_{P_j} = 1$. Replacing P_j with $V(\lambda_N, 1)$ will produce the same post measurement state as postselecting P_j on outcome $+1$. If a dependent measurement's determined outcome (as in (c)(i)) is inconsistent with an imposed postselection at that stage, then this indicates that the postselection requirement of the original circuit had probability zero. This results in a PBC process, some of whose measurements (arising from (c)(ii)) may still be postselected, completing the proof of Theorem 2.2.3(iii).

2.2.2 Proof of the extended Gottesman-Knill theorem

A PBC circuit with general input state ρ is similar to an adaptive Clifford circuit albeit with no unitary gate steps, except that the measurements are general Pauli measurements rather than just elementary Z measurements. Correspondingly our extended Gottesman-Knill Theorem 2.2.1 is obtained as a translation of Theorem 2.2.3 into a standard circuit form.

Proof of Theorem 2.2.1
According to Theorem 2.2.3(i), $\mathcal{C}$ can be weakly simulated by a PBC circuit of Pauli measurements $\tilde{P}_1, ..., \tilde{P}_s$ on input state ρ, and we just need to translate this back into an adaptive Clifford circuit with only Z basis measurements. This follows immediately by applying lemma 2.2.6 below to each $\tilde{P}_i$ separately, expressing it as $\tilde{P}_i = U_i^\dagger Z_k U_i$ for unitary Clifford operations U_i and any choice of line k (which could even be independent of i), thus establishing (i) and (iii).
Note that the Lemma cannot be applied to all $\tilde{P}_i$ simultaneously (giving a single U) since although pairwise commuting and independent, they are generally adaptively determined and not fixed a priori. However if $\mathcal{C}$ is non-adaptive then according to Theorem 2.2.3(ii), the sequence $\tilde{P}_1, ..., \tilde{P}_s$ can be chosen to be non-adaptive. Lemma 2.2.6 can then be applied to the whole list to give a single U with $U^\dagger Z_k U = \tilde{P}_k$ for $k = 1, \ldots, s$. The circuit $\mathcal{C}^*$ is then just the unitary Clifford

U (as unitaries after the Z measurements have no effect and can be deleted), thus establishing (ii).

Lemma 2.2.6. *Let $\{P_1,...,P_m\}$ be any set of independent and pairwise commuting Pauli operations on n qubits (so $m \leq n$). Then there is a unitary Clifford operation U such that $U^\dagger Z_k U = P_k$ for $k = 1,\ldots,m$. Furthermore a circuit of basic Clifford gates of depth $O(n^2/log(n))$ implementing U may be determined in classical poly(n) time.*

Proof. We first extend the set $\{P_1,...,P_m\}$ to a maximally sized set $\{P_1,...,P_n\}$ of independent pairwise commuting Pauli operations. This extension is not unique, but see Section 7.9 of [Preskill] for an efficient method of extension. Using similar techniques we also find generators of the 'destabiliser group' $\{D_1,...,D_n\}$ (defined in [5]). Then there is a unique (up to phase) Clifford V such that $VZ_iV^\dagger = P_i$ and $VX_iV^\dagger = D_i$ for $i = 1,\ldots,n$. An $O(n^2/log(n))$ circuit implementing V may be determined in classical poly(n) time by the construction of Theorem 8 in [5]. Finally take $U = V^\dagger$. □

2.3 Remarks and open questions

In this chapter we showed that, in the case of Clifford circuits, the 'classical' part of the circuit can be separated from the 'quantum'. This allowed us to classically simulate the latter and quantumly simulate the rest. The obvious open question is whether this sort of separation is possible for other classically simulable systems; if a class is classically simulable but becomes universal with the addition of an extra resource, is it possible to simulate the combination of these using a hybrid quantum–classical simulation?

One case we were particularly interested in was that of matchgates (discussed in Section 1.3). Matchgates are also known to have their own set of magic states. Allowing these magic states to the input of a matchgate circuit make it universal for computation. Is a (universal) matchgate circuit with input $|\psi_G\rangle \otimes |\psi_m\rangle$, where $|\psi_m\rangle$ is a magic state, simulable using 1) a matchgate circuit with input $|\psi_m\rangle$, 2) efficient classical processing? This would be the analogous result to the Extended Gottesman–Knill theorem. However, the proof of our theorem does not seem to easily (or effortfully) adapt to the matchgate case. It would appear that even though matchgates are very similar to Cliffords in many ways, proving this analogous theorem will require new techniques. Nevertheless, we hypothesis that this result is true.

Chapter 3

Quantum algorithms that cannot be simulated

A fundamental goal of quantum complexity theory is to prove that quantum computers cannot be efficiently simulated by classical computers. An approach to proving this was put forward by Bremner et al. [24], showing that if a particular class of quantum circuits, so-called IQP circuits, could be efficiently classically simulated up to multiplicative error then the polynomial hierarchy (PH) would collapse. However on physical grounds it is more natural to consider classical simulations with additive or l_1 error. In this vein, Aaronson and Arkhipov [3] showed that assuming the validity of two plausible complexity theoretic conjectures, the quantum process of boson sampling cannot be efficiently simulated up to additive error unless there is PH collapse. The conjectures are referred to as the anticoncentration conjecture and average-case hardness conjecture. Bremner, Montanaro and Shepherd [26] showed a similar result for IQP circuits, and furthermore they were able to prove the anticoncentration conjecture in their context. Since then, there have been further similar results for various classes [62, 15, 70, 61, 57, 44, 68].

In this chapter we introduce a subclass of quantum computing that we call Clifford Magic (CM) and establish a variety of its properties.The class CM comprises quantum circuits of unitary Clifford gates with fixed input $|A\rangle^{\otimes t}$ (for t qubit lines) and with output given by final measurement of some number of qubits in the computational basis. For computational applications we will also allow classical polynomial time computation for assistance before and after the Clifford circuit is run, in particular to determine the structure of a CM process $\mathcal{C}_w$ for each computational input bit string w. If the Clifford gates could adaptively depend on further intermediate measurements (not allowed here), the latter model would be universal for quantum computation, but our model appears to be weaker than universal. Our main result is to show that nevertheless, this class is hard to classically simulate up to additive error, given any one of a broad variety of average-case hardness conjectures.

This result is established by using the (quantum) simulation described in the Extended Gottesman–Knill theorem (Theorem 2.2.1). This theorem allows us to use CM circuits to emulate other classes of circuits that are known to be hard to classically simulate. For example CM can simulate certain IQP circuits. Therefore the same conjectures that, if proven, would show that the original IQP circuits are hard to simulate would show the same for CM is also.

This result has been shown in the recent works [15] and [70] (and our results were developed independently concurrently) but only for a single particular hardness conjecture. Furthermore both papers prove the anticoncentration conjecture by using the fact that random Clifford circuits form a k-design for suitable k. The idea of using k-designs to prove anticoncentration conjectures is explored in [45]. The reason our result provides stronger evidence that Clifford Magic is hard to simulate is then two fold. Firstly, we provide multiple conjectures such that if any one of these is proven the result follows. In contrast the other papers only give one conjecture. Secondly, our conjectures involve natural problems in $\#P$ for which the average case hardness result is more likely to hold. For example, because CM can simulate some IQP circuits, one of the conjectures is about the Ising Model (this will be explained in detail below). In contrast, the conjectures in the other papers about CM are not based on natural problems and therefore there is less reason to believe the average case hardness conjecture for those cases.

In the chapter we also describe potential experimental advantages to implementing CM circuits as opposed to some other schemes for experimental verification of quantum supremacy.

3.1 Quantum supremacy arguments

In this section we will recap the main method for showing 'quantum supremacy', the property of a class of quantum circuits that shows they are cannot be classically simulated unless plausible conjectures are false. We will particularly focus on the case where the classical simulations are required to accurate to additive error.

3.1.1 Additive case

The method in the following subsection will follow the arguments of Ref [3] and in particular Ref [26] but with some generalisation of context for our later purposes. This general method has been used many times in the literature (for example in [3, 26, Fefferman and Umans, 62, 15]) to argue for hardness of classical simulation, up to additive error, of a variety of classes of quantum computational processes.

Consider a given class $\mathcal{C} = \{C_\theta : \theta \in \Theta\}$ of quantum circuits parameterised by $\theta \in \Theta$, with each circuit also having its input state specified. We will generically denote the number of qubit lines of C_θ by n. Let the output be given by a measurement of all n lines and let $p_\theta(x)$ with $x \in B_n$ denote the output probability distribution of C_θ.

Introduce the following computational (sampling) task $\mathcal{T}_\mathcal{C}$ associated to the class $\mathcal{C}$: for any given θ, return (θ, y) where $y \in B_n$ has been sampled according to the output distribution p_θ of C_θ. We will be interested in the complexity of simulating this task (and some approximate variants) as a function of n.

By an ε-additive error simulation of the task $\mathcal{T}_\mathcal{C}$, we mean a process that given θ, returns (θ, y') where y' has been sampled according to a distribution q_θ on B_n which is an ε-additive approximation of the distribution p_θ.

An alternative task (that neither a classical nor quantum computer is likely to be able to efficiently achieve) is to compute a value for $p_\theta(x)$ for given θ and x, up to a (suitably specified) multiplicative error. Indeed for relevant classes that are studied in the literature, it can be shown that computing such approximations is #P hard in the worst-case. This task is of computational significance since for suitably chosen classes $\mathcal{C}$ the probability values can be used to represent quantities that are of independent physical or mathematical interest.

Our aim is to argue for classical hardness of simulation of the sampling problem $\mathcal{T}_\mathcal{C}$ up to additive approximation. To do this we will need to conjecture that estimating the value of $p_\theta(x)$ up to (suitable) multiplicative approximation remains #P hard not just in the worst-case, but in an average-case setting of the following kind.

For each class $\mathcal{C}$ and number of lines m introduce the set

$$\mathcal{D} = \{(\theta, x) : C_\theta \text{ has } m \text{ lines and } x \in B_m\}.$$

For each m we have a given probability measure π on the set of θ's that occur in $\mathcal{D}$, and let ν denote the uniform probability measure on B_m. Then $\pi \times \nu$ is the product measure on $\mathcal{D}$. Finally, to the class $\mathcal{C}$ we associate two constants: a measure size $0 < f < 1$ and an error tolerance η. We introduce the following conjecture that we will refer to as Hardness$(\mathcal{C}, \pi)$.

Average-case hardness conjecture for $\mathcal{C}$ with π: *let $\mathcal{F} \subseteq \mathcal{D}$ be any chosen subset of $\mathcal{D}$ having $\pi \times \nu$ probability measure f. Then it is #P hard to approximate the values $p_\theta(x)$ for all $(\theta, x) \in \mathcal{F}$ up to multiplicative error η.*

Note that if π is the uniform measure too, then the subsets $\mathcal{F}$ (for each m) will also be of

fractional size f. But for nonuniform π's there will be subsets of measure f that have smaller fractional size than f and asserting their #P hardness is a stronger conjecture. The use of nonuniform distributions will also feature significantly in the anticoncentration property below. As an example, in [26] classes of IQP circuits C are considered and conjectures 2 and 3 of [26] can be expressed as above, with π being the uniform distribution, $f = 1/24$ and $\eta = 1/4 + o(1)$. In [25] the authors also consider the same classes of IQP circuits, but a nonuniform π is used. This leads to a different average case hardness conjecture from those appearing in [26].

The arguments below will use several complexity classes that we will loosely describe here in a way that suffices to express the hardness of simulation argument. For more complete descriptions see for example Ref[8]. BPP^{NP} is the class of decision problems that can be solved by randomised classical polynomial time computations armed with an oracle for any problem in NP. FBPP^{NP} is the same except that the outputs can be bit strings rather than just a single bit. BPP^{NP} is in the third level of the tower of complexity classes known as the polynomial hierarchy PH. $\text{P}^{\#\text{P}}$ is the class of decision problems solvable in classical polynomial time, given access to an oracle for any #P problem; and it is known (Toda's theorem) that $\text{PH} \subseteq \text{P}^{\#\text{P}}$.

Now suppose that the sampling task $\mathcal{T}_{\mathcal{C}}$ can be solved up to additive error by a classical polynomial time algorithm $\mathcal{A}$. The first step is to show this ability to sample implies the existence of an FBPP^{NP} algorithm which, with use of $\mathcal{A}$, can estimate $p_\theta(x)$ up to an additive error, for each θ and a constant fraction of choices of x. After that an anticoncentration result will be used to convert the additive error into a multiplicative one, at least for a good measure of instances of (θ, x). The final step is to then invoke the average-case hardness conjecture for $\mathcal{C}$: if our multiplicative approximation determination (computable in FBPP^{NP}) is #P hard then $\text{P}^{\#\text{P}} \subseteq \text{P}^{\text{FBPP}^{\text{NP}}} = \text{BPP}^{\text{NP}}$. The latter class is in the third level of PH and then by Toda's theorem, PH will collapse to its third level. However such a collapse is widely regarded as extremely implausible (similar to a collapse of NP to P), providing plausibility that the purported classical polynomial time algorithm $\mathcal{A}$ for solving $\mathcal{T}_{\mathcal{C}}$ up to additive error, cannot exist (if the average hardness conjecture is accepted).

Lemma 3.1.1. *(adapted from Lemma 4 of [26]) Suppose there is a classical polynomial time algorithm $\mathcal{A}$ that simulates the sampling task $\mathcal{T}_{\mathcal{C}}$ up to additive error ε. Then for any $0 < \delta < 1$ there is an FBPP^{NP} algorithm that, for each θ, approximates $p_\theta(x)$ up to additive error*

$$\frac{p_\theta(x)}{poly(n)} + (1 + o(1)) \cdot \frac{\varepsilon}{2^n \delta} \tag{3.1}$$

for at least a fraction $1 - \delta$ of all $x \in B_n$. Thus for any probability measure π, the subset of $\mathcal{D}$ to which eq. (3.1) applies, has $\pi \times \nu$ measure at least $1 - \delta$ (since the measure of the full space of θ's is always unity).

This lemma is readily proved by following the argument of the proof of Lemma 4 in [26], with minor notational modifications.
To obtain a multiplicative error from this additive one, we require an anticoncentration property of the following form.

Anticoncentration property for $\mathcal{C}$ with π: *there are constants $\alpha > 0$ and $0 \leq \beta \leq 1$ such that $p_\theta(x) \geq \alpha/2^n$ holds on a subset of $\mathcal{D}$ of $\pi \times \nu$ measure at least β.*

In the literature a property of this form is proved for some classes $\mathcal{C}$ (e.g. in [26, 15, 62, 25]) and conjectured to hold for others (e.g. in [3]). Proofs of the property generally involve applying the Paley-Zygmund inequality to the probability measure $\pi \times \nu$.
Suppose now that the anticoncentration property holds for $\mathcal{C}$. Then by choosing δ in Lemma 3.1.1 to be $\beta/2$ we guarantee an overlap $\Xi \subset \mathcal{D}$ of probability measure at least $\beta/2$ on which the anticoncentration property $p_\theta(x)/\alpha \geq 1/2^n$ and the additive approximation bound of eq. (3.1) both hold.

Then substituting $p_\theta(x)/\alpha$ for $1/2^n$ in eq. (3.1) the approximation bound becomes

$$\frac{p_\theta(x)}{\text{poly}(n)} + (1 + o(1)) \cdot \frac{2\varepsilon}{\alpha\beta} p_\theta(x)$$

giving a multiplicative approximation bound of size $\frac{2\varepsilon}{\alpha\beta} + o(1)$ for $p_\theta(x)$, for a $\beta/2$ measure subset of $\mathcal{D}$.
Finally collecting all the above, we arrive at the following conclusion.

Theorem 3.1.2. *Let $\mathcal{C}$ be any class of quantum circuits with associated measure π for which the anticoncentration property holds (with constants α and β). Suppose that the sampling task $\mathcal{T}_\mathcal{C}$ can be efficiently classically simulated up to additive error ε. Then if the average-case hardness conjecture holds with measure size $f = \beta/2$ and error tolerance $\eta = 2\varepsilon/(\alpha\beta)$, the polynomial hierarchy will collapse to its third level.*

For example in [26] we have $\varepsilon = 1/192$, and the anticoncentration property is shown to hold with uniform π, $\alpha = 1/2$ and $\beta = 1/12$. So to obtain collapse of PH we need the average-case hardness conjecture to be valid with error tolerance $\eta = 2\varepsilon/(\alpha\beta) = 1/4$ and fraction $f = \beta/2 = 1/24$.

3.2 Quantum advantage of Clifford Magic circuits

3.3 Clifford magic (CM) circuits

We introduce a class of quantum processes that we call “Clifford Magic", written CM.

Definition 3.3.1. *A CM circuit on t qubits is a unitary Clifford circuit which has input state $|A\rangle^{\otimes t}$, and output given by the result of measuring r specified qubits (the output register $\mathcal{O}$) in the Z basis (and intermediate measurements are not allowed). A postselected CM circuit is a CM circuit with an additional register $\mathcal{P}$ of s qubits (called the postselection register) disjoint from $\mathcal{O}$, which is also measured at the end.*

Our motivation for introducing and studying CM circuits is twofold. The first reason, discussed in Subsection 3.3.1, relates CM processes to known classical simulation results. In particular, we show that the class of CM circuits is equivalent to a class of quantum circuits likely to have supra-classical power while also being weaker than BQP. Our second motivation, discussed in Subsection 3.5, is that CM circuits are a promising candidate for experimentally verifying quantum advantage. Unlike other quantum supremacy proposals, small amounts of error correction can be readily included with modest overheads. Furthermore, adding adaptive measurements to CM processes makes the class universal while also providing an economy in the number of qubits needed, as described previously in Figure 1. In this way CM circuits may be viewed as a practicable stepping stone towards an implementation of universal quantum computation.

3.3.1 Relation between CM and known classical simulation results

Consider circuits of the form shown in Figure 2. The circuits on the left comprise unitary Clifford gates with input $|0\rangle^{\otimes n}|A\rangle^{\otimes \mathrm{poly}(n)}$ and one line being measured for the output. Such circuits are known to be classically simulable [53]. On the other hand, if intermediate Z measurements are allowed together with adaptations, the circuits can perform T-gadgets making them universal for BQP computations, as shown on the right.
Consider now the family of all Clifford circuits with input $|0\rangle^{\otimes n}|A\rangle^{\otimes \mathrm{poly}(n)}$ and one line being measured for the final output, and allowing intermediate measurements. Let $\mathcal{M}_I$ denote the set of intermediate measurement results obtained. Then we can consider $\mathcal{M}_I$ being used in one of the following three ways:
(A) Discarding $\mathcal{M}_I$, and not using it in any way (either for output or for adaptations).
(B) Retaining $\mathcal{M}_I$ as part of the output (but not used otherwise).
(C) Using $\mathcal{M}_I$ as it emerges for subsequent adaptation in the course of the process, as well as giving $\mathcal{M}_I$ as part of the output.

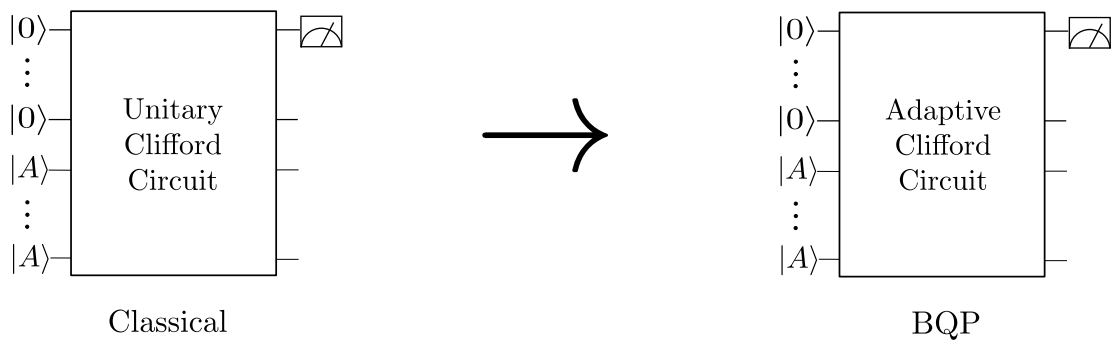


Figure 3.1 The circuits on the left have magic states as well as stabiliser inputs. However, if a unitary Clifford circuit is applied and only one line is measured, it is classically simulable. On the other hand, if intermediate Z measurements are included and the circuit is allowed to adaptively depend on measurement outcomes, then the circuit can perform any BQP computation.

Circuits of the form (C) can perform any BQP computation, but those of the form (A) are classically simulable [53]. Case (B) is not expected to have the full power of BQP. But furthermore, using the methods of [53] (cf especially Theorems 6 and 7 therein, and under plausible complexity conjectures) case (B) is also not classically simulable exactly (in either the strong or weak sense). In this work (cf Section 3.4) we will show that additionally, it is also not classically simulable up to multiplicative or additive error either (under plausible conjectures).

Case (B) is clearly intermediate between (A) and (C). Indeed (C) allows the extra capability over (B) of adaptation, and compared to (A), retaining $\mathcal{M}_I$ in (B) gives more information about the final state which in (A) would be assigned as the probabilistic mixture of all post-measurement states arising from all the possible outcome values for $\mathcal{M}_I$.

The class of CM circuits is clearly a subset of the class of circuits in case (B) viz. those with no $|0\rangle$ part in the input and all measurements being performed only at the end. However, the CM subset is in fact equivalent to the full class in (B): every circuit in the latter can be weakly simulated by a CM circuit, as follows by an application of the Extended Gottesman–Knill theorem. As the intermediate measurements in case (B) are not adaptive, Theorem 2.2.1(ii) tells us that the resulting compressed circuit is a CM circuit.

In this sense the computational power of the class of CM circuits relates directly to the power of retaining intermediate measurements in a Clifford circuit. We prove in Section 3.4 that CM circuits cannot be classically simulated (up to multiplicative or additive error) under plausible conjectures, showing that the mere retention of intermediate measurement results as above, can be regarded as a kind of "quantum resource", elevating the classically simulable case (A) to

supra-classical computing power in (B).

3.4 Hardness of classical simulation of CM circuits

We now establish lower bounds on the complexity of classical simulation of CM circuits, allowing either multiplicative or additive errors in the simulation. The scenario of additive error is generally regarded as a reasonable model of what is feasible to physically implement in practice.
A distribution $q(x)$ is an ε-additive approximation of a distribution $p(x)$ if

$$\sum_x |p(x) - q(x)| \leq \varepsilon. \tag{3.2}$$

A number Y is an ε-multiplicative approximation of a number X if $|X - Y| \leq \varepsilon X$. A distribution $q(x)$ is an ε-multiplicative approximation of a distribution $p(x)$ if for each x, $q(x)$ is an ε-multiplicative approximation of $p(x)$. Thus clearly ε-multiplicative approximation of distributions implies ε-additive approximation.

3.4.1 Hardness of classical simulation of CM with multiplicative error

Although (uniform families of) CM circuits themselves are not likely to be universal for quantum computation, we first establish that postselected CM circuits suffice as a quantum resource for postselected universal quantum computation. Using the arguments of Ref [24], this is enough to establish that the class cannot be classically simulated to multiplicative error without causing the Polynomial Hierarchy (PH) to collapse.

Theorem 3.4.1. *Any postselected poly-sized unitary quantum circuit $\mathcal{C}$ on n qubits (with final Z measurements) can be weakly simulated by a postselected poly-sized CM circuit on poly(n) qubits.*

Proof. We may suppose without loss of generality that $\mathcal{C}$ has the following form: the input state is $|0\rangle^{\otimes n}$, followed by Clifford and T gates, and finally some number of lines is measured in the Z basis. Of these, some are postselected to outcome $k = +1$. To begin, we replace each T gate with a T-gadget where the gadget measurement is postselected to outcome $+1$ so the correction S is not required. As no other part of the circuit acts on this ancilla line again this measurement can be performed at the end of the circuit. The resulting circuit $\tilde{\mathcal{C}}$ then has input $|0\rangle^{\otimes n}|A\rangle^{\otimes t}$, which is acted on by a Clifford unitary U followed by Z measurements, some of which are postselected. The proof is now completed in either one of two possible ways, labelled (a) and (b), as follows:
(a) Theorem 2.2.1(ii) and (iii) can then be used to provide an algorithm for simulating the above

circuit $\tilde{C}$ by a postselected CM circuit.
(b) We start with the state $|A\rangle^{\otimes(n+t)}$ and first convert it to $|0\rangle^{\otimes n}|A\rangle^{\otimes t}$. This is achieved by applying a T-gadget postselected to outcome -1 (thus implementing a $T^\dagger$ gate), and then H, to each of the first n qubits, and then we apply the Clifford unitary U and final Z measurements above. As the gadget measurements can be moved to the end, this whole process is a postselected CM circuit.

□

Corollary 3.4.2. *Any language in post-BQP can be decided with bounded error by a postselected CM circuit assisted by efficient classical computation. Thus if uniform families of CM circuits could be weakly classically simulated to within multiplicative error $1 \leq c < \sqrt{2}$, then the polynomial hierarchy would collapse to its third level.*

Proof. The first claim follows immediately from Theorem 3.4.1, and then the second follows from [24]. □

3.4.2 Hardness of classical simulation of CM with additive error

We now show that CM circuits cannot be classically efficiently simulated with additive error unless PH collapses, given average-case hardness conjectures. While CM circuits have been shown before [15, 70] to have this property for one particular average-case-conjecture, here we show that actually a broad variety of such conjectures apply, such that if any one of them is proven, it implies the hardness of CM circuit simulation. Furthermore, in previous work, this hardness result for CM was shown by invoking the fact that Clifford gates form a 2-design [31] and that 2-designs anticoncentrate [45, 57], to give the needed anticoncentration property. Here we follow a very different method, instead using the ability of CM circuits (via Therorem 2.2.1) to simulate any nonadaptive circuit. This allows CM circuits to simulate several other classes of circuits (not necessarily 2-designs) and inherit their average-case hardness conjecture as a basis for hardness of CM circuit simulation up to additive error.

Consider any class of unitary circuits $\mathcal{C} = \{C_\theta : \theta \in \Theta\}$ and associated measure π on Θ, for which a suitable anticoncentration property holds, and whose classical simulation up to additive error would imply collapse of PH if we assume Hardness$(\mathcal{C}, \pi)$. Suppose that these circuits have been expressed as circuits of gates from the universal set of basic Clifford gates with T and $T^\dagger$. We can use any choice of such a representation. Now consider the expanded class $\mathcal{C}^T$ obtained by taking each circuit C_θ and replacing each T and $T^\dagger$ gate by either T or $T^\dagger$ in all combinations. If C_θ has t T and $T^\dagger$ gates then it will give rise to 2^t circuits in $\mathcal{C}^T$, and these can be labelled by (θ, τ) where τ is a t-bit string indicating the choices of T and $T^\dagger$. Accordingly, we write $\mathcal{C}^T = \{C_{\theta,\tau} : \theta \in \Theta, \tau \in B_t\}$.

$\mathcal{C}^T$ is exactly the class of circuits we obtain if we implement the circuits $\mathcal{C}_\theta$ using T gadgets for each T and $T^\dagger$ gate, but omit all the adaptive S gate corrections that are normally specified by the T-gadget measurement outcomes. Denote that non-adaptive circuit by U_θ with outputs (x,τ) where $\tau \in B_t$ is the string of gadget measurement outcomes and x arises from the output lines from C_θ. Each of the 2^t possibilities for τ will occur with equal probability. Note that the circuits U_θ are unitary Clifford circuits (having only final Z measurements). Indeed the measurement within any (generally intermediate) T-gadget can now be moved to the end of the circuit as that line is not acted on again, and the measurement outcome is not used in any adaptations. Because these circuits are unitary Clifford circuits, they can be simulated by CM circuits using Theorem 2.2.1 (ii). Denote the associated CM circuit (with input state $|A\rangle^{\otimes t}$) by V_θ. Finally let $p_\theta(x)$, $p_{\theta,\tau}(x)$ and $u_\theta(x,\tau)$ (with $x \in B_n$, $\tau \in B_t$) denote the output probabilities for the circuits C_θ, $C_{\theta,\tau}$ and U_θ respectively.

Note that for each θ there is a $\tau_0 = \tau_0(\theta)$ for which $p_{\theta,\tau_0}(x) = p_\theta(x)$, viz. τ_0 just specifies the T and $T^\dagger$ choices that actually occur in C_θ. Furthermore, since each τ arises in the output of U_θ with equal probability $1/2^t$, the relationship between $C_{\theta,\tau}$ and U_θ gives (via conditional probabilities):

$$p_{\theta,\tau}(x) = u_\theta(x,\tau)\,2^t. \tag{3.3}$$

Finally in addition to distribution π on the θ's, let ν and ν' denote the uniform distribution on the x's and τ's respectively. Let $\mathrm{prob}_{\pi\times\nu\times\nu'}(\theta,x,\tau)$ denote the probability of (θ,x,τ) in the product distribution $\pi\times\nu\times\nu'$, and similarly for $\mathrm{prob}_{\pi\times\nu'}(\theta,\tau)$, $\mathrm{prob}_\pi(\theta)$ etc.
We will show that, for some classes $\mathcal{C}$ of circuits already proved to have the additive simulation hardness property of Theorem 3.1.2 (subject to an associated Hardness$(\mathcal{C},\pi)$ conjecture), that $\mathcal{C}^T$ contains no new circuits that were not already present in $\mathcal{C}$. Thus the labels (θ,τ) will label the circuits of $\mathcal{C}$ with generally high redundancy, and we write $\mathcal{C}^T = \mathcal{C}$ in this situation. Since such circuits can be simulated by CM circuits, classical simulation of CM circuits up to additive error can then imply collapse of PH, subject to the conjecture Hardness$(\mathcal{C},\pi)$ of the class $\mathcal{C}$, as will be formalised in the Theorem below.

Suppose now that $\mathcal{C} = \mathcal{C}^T$. Then for each (θ,τ) there is $\tilde{\theta} = \tilde{\theta}(\theta,\tau)$ with $C_{\theta,\tau}$ being $C_{\tilde{\theta}}$ so

$$p_{\theta,\tau}(x) = p_{\tilde{\theta}}(x).$$

We will also require the following θ-sampling relation: the C_θ circuits occurring multiply in $\mathcal{C}^T$, occur with the same probability in $\mathcal{C}^T$ (wrt distribution $\pi\times\nu'$) as they did in $\mathcal{C}$ (wrt distribution π):

$$\sum_{(\theta,\tau):\tilde{\theta}(\theta,\tau)=\theta_0} \mathrm{prob}_{\pi\times\nu'}(\theta,\tau) = \mathrm{prob}_\pi(\theta_0). \tag{3.4}$$

Theorem 3.4.3. *Consider any class of circuits $\mathcal{C}$ with associated distribution π for which the following hold:*
(i) the anticoncentration property (with parameters α and β);
(ii) $\mathcal{C} = \mathcal{C}^T$ and the θ-sampling relation eq. (3.4).
Then if every CM circuit can be efficiently classically simulated to additive error ε, the average-case hardness conjecture for $(\mathcal{C}, \pi)$ with parameters $f = \beta/2$ and $\eta = 2\varepsilon/(\alpha\beta)$ will imply that PH collapses.

Proof. We use the notations and definitions introduced above. Since U_θ can be simulated by a CM circuit, if every CM circuit can be efficiently classically simulated to additive error ε, then so can the distribution $u_\theta(x,\tau)$. So by Lemma 3.1.1 applied in (θ,τ,x) space, there is a $(1-\beta/2)$ sized subset in $\pi \times \nu' \times \nu$ measure where an FBPP$^{\text{NP}}$ algorithm can calculate an additive approximation to $u_\theta(x,\tau)$ with additive error bound of

$$\frac{u_\theta(x,\tau)}{\text{poly}(n+t)} + (1+o(1)) \cdot \frac{2\varepsilon}{2^{n+t}\beta} \tag{3.5}$$

(since we have $n+t$ lines now).
Next we will want a measure β subset of (θ,τ,x)'s on which the anticoncentration property $u_\theta(\tau,x) \geq \alpha/2^{n+t}$ holds. By $(\mathcal{C},\pi)$ anticoncentration, there is a measure β subset of (θ,x)'s with $p_\theta(x) \geq \alpha/2^n$. So by the θ-sampling relation eq. (3.4) and eq. (3.3) there is a measure β subset of (θ,τ,x)'s with

$$u_\theta(x,\tau) = \frac{p_{\theta,\tau}(x)}{2^t} \geq \frac{\alpha}{2^{n+t}} \tag{3.6}$$

(noting that for any x, $\text{prob}_{\pi\times\nu}(\theta,x) = \text{prob}_\pi(\theta)/2^n$). Combining eqs. (3.6) and (3.5) we get a measure $\beta/2$ subset of (θ,τ,x)'s on which $u_\theta(x,\tau)$ can be calculated by an FBPP$^{\text{NP}}$ algorithm to multiplicative approximation $2\varepsilon/(\alpha\beta)+o(1)$, and this also applies to $p_{\theta,\tau}(x) = u_\theta(x,\tau)2^t$ (as multiplicative approximations are invariant under scale changes).
Finally we want to map this back to (θ,x) space. Note that for any x

$$\text{prob}_{\pi\times\nu'\times\nu}(\theta,\tau,x) = \frac{1}{2^n}\text{prob}_{\pi\times\nu'}(\theta,\tau) \leq \frac{1}{2^n}\text{prob}_\pi(\tilde{\theta}(\theta,\tau)) = \text{prob}_{\pi\times\nu}(\tilde{\theta},x)$$

(where the inequality follows from eq. (3.4)). Hence the map $(\theta,\tau,x) \mapsto (\tilde{\theta}(\theta,\tau),x)$ gives a subset of (θ,x)'s of measure $\geq \beta/2$ on which $p_\theta(x)$ can be calculated to multiplicative approximation $2\varepsilon/(\alpha\beta)+o(1)$ by an FBPP$^{\text{NP}}$ algorithm. Hence the average-case hardness conjecture for $(\mathcal{C},\pi)$ implies that PH collapses to its third level. □

Examples of circuit classes in the literature for which a suitable anticoncentration property holds, $\mathcal{C} = \mathcal{C}^T$ and the θ-sampling relation eq. (3.4) holds, include the following.

IQP circuits associated with the Ising model [26]
This is the class of circuits $\mathcal{C}$ having input $|0\rangle^{\otimes n}$ acted on by $H^{\otimes n}UH^{\otimes n}$, where U is unitary and

chosen in the following way: apply T^{v_i} to each qubit line i, and $CS^{w_{ij}}$ to each pair of qubits i, j, where v_i and w_{ij} (all collectively comprising the label θ) are chosen in all possible combinations from $\{0,...,7\}$ and $\{0,...,3\}$ respectively, and CS is the controlled-S gate. Furthermore the CS gate is implemented in terms of Clifford+T+T† gates using the gadget of Figure 3. The distribution π is the uniform distribution.

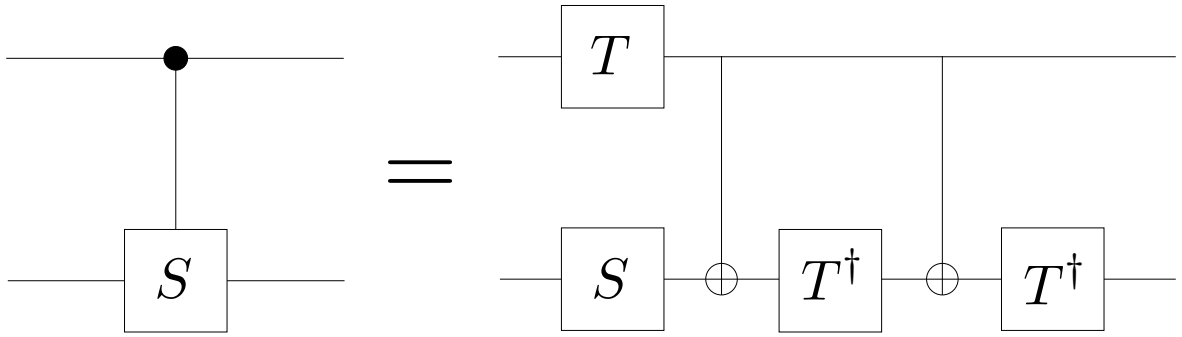


Figure 3.2 Decomposing the controlled-S gate into Clifford+T+$T^\dagger$ gates.

To see that $\mathcal{C} = \mathcal{C}^T$ note first that if any initial T or $T^\dagger$ gates are changed (to the other choice), the resulting circuit is clearly still a circuit in the original set. However, there are also T and $T^\dagger$ gates within the CS gadget of Figure 3 to consider. If the T or $T^\dagger$ gates at either end are changed, this can be corrected by applying further T gates. If the middle $T^\dagger$ gate is swapped, the result is $CS(T \otimes T)$. So in each of these cases, the resulting circuit is still from the original set. The θ-sampling relation eq. (3.4) holds because for each θ there is a $\tau_0 = \tau_0(\theta)$ with $\tilde{\theta}(\theta, \tau_0) = \theta$ and the fact that for any fixed τ' (and varying θ) the mapping $(\theta, \tau_0(\theta)) \mapsto (\theta, \tau_0 \oplus \tau')$ is one-to-one on the underlying $\tilde{\theta}$'s (with $\oplus$ being addition of t-bits strings at each entry).

Sparse IQP circuits [25]

This class is the same as the above (so $\mathcal{C} = \mathcal{C}^T$) but with a different distribution π. Specifically, having chosen each v_i and w_{ij} uniformly, each $CS^{w_{ij}}$ gate is applied only with some probability p, while each T^{v_i} is applied as in the above case. This amounts to $w_{ij} = 0$ being chosen with probability $\frac{1}{4} + \frac{3}{4}(1-p)$ and other w_{ij}'s with probability $p/4$ (and v_i's chosen uniformly as before). Also as before when a T gate inside of CS is swapped, it always becomes CS with some extra T gates. The θ-sampling relation eq. (3.4) holds since reassigning T and $T^\dagger$ gates always preserves the number of two qubit gates in the circuit.

Random Circuit Sampling [13]

Another class of circuits was put forward by the Google/UCSB team, and called random circuit sampling. The gates used in these circuits are from $\{CZ, X^{1/2}, Y^{1/2}, T\}$. In [45] it is shown that circuits from this set anticoncentrate if they are chosen as follows: let $G = \{CZ, X^{1/2}, X^{-1/2}, Y^{1/2}, Y^{-1/2}, T, T^\dagger\}$ (i.e. the previous set closed under inverses). In each time

step either $U_{1,2} \otimes U_{3,4} \otimes ... \otimes U_{n-1,n}$ or $U_{2,3} \otimes U_{4,5} \otimes ... \otimes U_{n-2,n-1}$ is applied, for all possible choices of $U_{j,j+1}$ from G (with 1-qubit gates U appearing as $I \otimes U$ or $U \otimes I$). Finally all n lines are measured in the computational basis. The distribution π over $\mathcal{C}$ is the uniform distribution. All gates in G besides T and $T^\dagger$ are Clifford, so reassigning T and $T^\dagger$ gates clearly results in circuits from the same class i.e. $\mathcal{C} = \mathcal{C}^T$, and a uniform distribution for π satisfies eq. (3.4).
In [14] it is shown that Random Circuit Sampling has a property similar to the required average-case hardness result viz. that the conjecture holds if the task is to compute $p_\theta(x)$ exactly. This is known to be #P hard, even for the average case. Boson sampling [3] is the only other class where this is kind of result has been proved. Although referring to exact calculation, this can nevertheless be viewed as providing evidence that the necessary average-case hardness conjecture (involving approximate computation, up to multiplicative error) may hold.
CM circuits simulating any one of these three classes inherit the hardness of the original circuits. If average-case hardness is shown for any of them then it implies the same is true for CM circuits and therefore that CM cannot be efficiently classically simulated up to additive error. This result is a natural consequence of the Extended Gottesman–Knill theorem that shows how CM circuits can simulate other types of quantum computations.

For other classes of circuits we generally have $\mathcal{C} \neq \mathcal{C}^T$ i.e. $\mathcal{C}^T$ contains circuits that were not already present in $\mathcal{C}$. However, if $\mathcal{C}^T$ also has a suitable anticoncentration property, then up to an average-case hardness conjecture, PH will collapses if $\mathcal{C}^T$ circuits can be classically simulated to additive error. Note that if $\mathcal{C}$ has a worst-case hardness result (as is generally the case for classes considered), then so does $\mathcal{C}^T$ since its circuits always form a superset of $\mathcal{C}$. This provides evidence for a suitably analogous average-case conjecture for $\mathcal{C}^T$. Hence, in the case that $\mathcal{C}^T$ also anticoncentrates, it is also likely to be hard to classically simulate. For any $\mathcal{C}$, the circuits in $\mathcal{C}^T$ can always be simulated by CM circuits (in the sense above, used in Theorem 3.4.3, taking the uniform distribution over the τ's as above) and we obtain the following result.

Theorem 3.4.4. *Suppose that $\mathcal{C}^T$ (arising from $(\mathcal{C}, \pi)$ as described above) satisfies an anti-concentration property with constants α and β. Then if every CM circuit can be efficiently classically simulated to additive error ε, PH will collapse to the third level if we assume an average hardness conjecture for $\mathcal{C}^T$ with parameters $f = \beta/2$ and $\eta = 2\varepsilon/(\alpha\beta)$, extending the corresponding conjecture for $\mathcal{C}$. Furthermore, if $\mathcal{C}$ had the worst-case hardness property, then so does $\mathcal{C}^T$.*

One example of circuits for which $\mathcal{C} \subsetneq \mathcal{C}^T$ and $\mathcal{C}^T$ also anticoncentrates, is the class of **Conjugated Clifford circuits** introduced in [15]. Here we have circuits of the form $V^{\otimes n\dagger} U V^{\otimes n}$, where V is any fixed 1-qubit gate and U is any Clifford circuit (so we get a class for each choice of V), and π is the uniform distribution. The representation of V in terms of Clifford+T+$T^\dagger$ gates generally contains T and $T^\dagger$ gates, and when these are reassigned in all combinations in $V^{\otimes n}$, the result is no longer necessarily a gate of the form $W^{\otimes n}$ i.e. the gates applied on different

lines will generally be different, and the n-qubit gate on one end will also not necessarily be the inverse of the one on the other end. Hence $\mathcal{C} \subsetneq \mathcal{C}^T$. However, this new class of circuits does anticoncentrate. This follows from the original anticoncentration proof in Ref [15] (Lemma 4.3 there) which still applies for arbitrary n-qubit gates replacing $V^{\otimes n}$ and $V^{\otimes n\dagger}$ on the ends.

3.5 Experimental advantages of CM circuits

CM circuits offer several advantages for fault tolerant implementation and for implementation in the MBQC model, inherited in part from such benefits for Clifford circuits.

Fault tolerance for CM circuits

In the circuit model, fault tolerance is often achieved by replacing T gates by T gadgets, with magic state distillation being used to create high fidelity $|A\rangle$ states offline [21]. However, as T gadgets include adaption, the circuit cannot be fully created in advance, and instead part of the circuit must be created in real time. These potentially increase the required coherence times. CM do not require these kinds of adaptions, even when made fault tolerant using a stabiliser code.

Syndrome measurements and their associated correction operations may appear to introduce further adaptations into the circuit, but these can in fact be avoided. Indeed these corrections are Pauli operations, and can always be commuted past Clifford unitaries and (Pauli) syndrome measurements, since the Pauli measurements, at most, swap sign when conjugated by the Pauli corrections. Then the Pauli corrections can be accounted for after the quantum computation is completed via simple classical processing of the measurement outcomes.
A further benefit of CM circuits being Clifford circuits is that any such circuit on t qubit lines can be expressed as a circuit of depth bounded by $O(t^2/\log t)$ [5], again providing potential benefits for shorter coherence times in implementation.

CM circuits in the MBQC model

In our discussion below we will assume the following standard form of MBQC (cf for example [32]). The starting resource state is the standard cluster state. CZ operations in circuits are implemented by exploiting CZ's that were used in the construction of the cluster state. 1-qubit measurements applied to the cluster state are either Z measurements or else $M(\alpha)$ measurements in the basis $\{|\pm_\alpha\rangle\}$, where $|\pm_\alpha\rangle = 1/\sqrt{2}(|0\rangle \pm e^{-i\alpha}|1\rangle)$. The latter provide implementation of 1-qubit gates $J(\alpha) = H(|0\rangle\langle 0| + e^{i\alpha}|1\rangle\langle 1|)$, appearing as $X^s J(\alpha)$ where $s = 0, 1$ is the measurement outcome and X^s is the associated byproduct operator. The $J(\alpha)$ gates together with CZ provide a universal set.

Theorem 3.5.1. *A CM circuit $\mathcal{C}$ including preparation of its input $|A\rangle^{\otimes t}$, can be implemented in the MBQC model in depth 1.*

Proof. Note first that $|A\rangle = HJ(\pi/4)|+\rangle$. Thus $\mathcal{C}$ may be viewed as having input $|+\rangle$ on all lines, followed by a round of $J(\pi/4)$ gates, followed by Clifford gates (comprising a round of H gates followed by the gates of $\mathcal{C}$). Hence for MBQC implementation the measurement pattern comprises a line of $M(\pi/4)$ measurements laid out next to implementations of Clifford gates. The X^s byproducts of the $M(\pi/4)$ measurements can be commuted over the Clifford gates to the end, without incurring any adaptations. Similarly it is well known [76] that Clifford circuits can be implemented without adaptation to the byproduct operators that arise. Hence the entire measurement pattern is non-adaptive and can be implemented in depth 1. □

Miller et al. [61] also propose a scheme for quantum supremacy without error correction that is depth 1 in MBQC, based on use of MBQC to simulate IQP circuits. Their scheme requires a nonstandard resource state that may not be simple to prepare, whereas our proposal uses the standard cluster state, which is a stabiliser state, as the resource. Furthermore our scheme can be made fault tolerant as follows.

Theorem 3.5.2. *A CM circuit $\mathcal{C}$ can be implemented fault tolerantly in the MBQC model in depth 1, given a particular initial resource state that can be created offline with high fidelity.*

Proof. For simplicity, we will consider a fault tolerance scheme using the 7-qubit Steane code. The initial resource state can be created as follows. Create an encoded magic state $|\tilde{A}\rangle^{\otimes t}$. Create the other parts of the encoded graph state by making the encoded states $|\tilde{+}\rangle$ and using the encoded version of CZ. The usual syndrome measurements and corrections are required during this process. Inclusion of $|\tilde{A}\rangle^{\otimes t}$ into the resource state allows us to avoid a later need for implementing encoded $M(\pi/4)$ measurements fault tolerantly, and our CM circuit is a circuit of only Clifford gates. Now we have $H = J(0)$ and $S = HJ(\pi/2)$, with $M(0)$ and $M(\pi/2)$ being X and Y measurements respectively. Thus in MBQC, Clifford gates are implemented using only Pauli measurements, and in our encoded setup we need to apply their corresponding fault tolerant encoded versions. These are transversal. Furthermore, syndrome measurements can be carried out using the usual fault tolerant construction in terms of Clifford operations and ancillas. These Clifford gates themselves can be implemented using MBQC using ancillas. All these ancillas are included in the initial state. Hence every physical operation applied to the initial state is a 1 qubit Pauli measurement. Then, as before, Pauli errors can be corrected via classical post processing, and so the circuit is depth 1. □

3.6 Remarks

The landmark of quantum supremacy was achieved in 2019 by experimenters from Google and University of California Santa Barbra [9]. The model they implemented was Random Circuit

Sampling [13]. Random Circuit Sampling makes use of universal gates, which makes it a good stepping stone toward practically useful quantum computation, a goal many are working toward.

Theoretical Quantum Supremacy results will continue to be important in the field of classical simulation. That is because they are a convenient tool for deciding whether a (restricted) quantum computer can perform truly quantum computations.

Chapter 4

Entanglement and the quantum computing advantage

What is the cause of the speedups that quantum computers can achieve? What quantum effects really give quantum computers their advantage? There are both practical and philosophical reasons to think about this question. The practical reason is that it will give a better idea of how to design fast quantum algorithms- an art that is still mysterious today. Philosophically, we want to know, what features of quantum mechanics cannot be simulated classically? This will give us clues about why quantum mechanics is such an unusual theory.

Bell's theorem identifies one feature of quantum mechanics that cannot be simulated [10]. It shows that without entanglement there cannot be violations of the so called "Bell inequalities" that are signatures of nonlocality. Therefore we see that entanglement is crucial to this sort of nonlocality that quantum mechanics displays[1].

Entanglement is also often conjectured to be the source of the speedups achieved by quantum computers. This claim is supported by the results of Jozsa and Linden [51] and Vidal [88], which showed that pure state computations with only small amounts of entanglement can be efficiently classically simulated. Therefore entanglement is clearly necessary for pure state quantum computation. However, it has not been shown that separable mixed state computations are classically simulable. Deciding this is one of the questions posed in the *"Ten Semi-Grand Challenges for Quantum Computing Theory"* raised by Aaronson in 2005 [1].

If mixed state quantum computers can obtain some advantage without entanglement, this would imply that entanglement is not the only source of quantum advantage. Such a result would be plausible even though it does not hold in the pure case. Pure states with restricted entanglement

[1]However, not all (mixed) entangled states violate Bell inequalities[90] suggesting that there is more to Bell violation than just entanglement.

can be described very efficiently, while for mixed states without entanglement this is not obviously the case.

One possible reason that entanglement may be necessary for pure, but not mixed, states is that entanglement is the only type of correlation present in the former, but not in the latter. Perhaps it is these other correlations that are important in a quantum computer, and not just entanglement itself. Indeed, Vidal's [88] algorithm can only efficiently simulate mixed state computations when the total correlations (entanglement as well as classical correlations) are restricted.

Another argument against entanglement being the only resource responsible for quantum advantages arises from studying the One Clean Qubit model, whose corresponding complexity class is known as DQC1 (Deterministic quantum computation with one clean qubit). This model is one of the few known nontrival mixed state quantum computers. The input state to this computer is one pure (or 'clean') qubit and n qubits in the maximally mixed state [55]. Any polynomial-sized quantum circuit can be applied to these qubits, after which the initially clean qubit is measured. Despite how unresourceful the initial state appears to be, DQC1 can perform several tasks seemingly exponentially faster than is classically possible, such as the following: computing the normalised trace of a unitary [79]; estimating a coefficient in the Pauli decomposition of a quantum circuit up to polynomial accuracy [55]; computing Schatten p-norms up to a suitable level of accuracy [30]; and computing the trace closure of a Jones polynomial[79]. The power of the class remains the same even if we allow more than one clean qubit [79] – that is, $\mathsf{DQC1} = \mathsf{DQCk}$ for $k = O(\log(n))$ – suggesting that the initial state is indeed more resourceful than it first appears to be. Furthermore, several results have shown that DQC1 cannot be classically simulated under some complexity theoretic conjectures [39, 64, 63].

This is all despite the fact that DQC1 does not require entanglement to be present between the clean qubit and the maximally mixed register in order to demonstrate an advantage over classical computation [74]. Moreover, the amount of entanglement in the computer, as measured by the multiplicative negativity, is always bounded by a constant independent of the number of qubits [33]. DQC1 can still have strong correlations besides entanglement [34], however, meaning that it cannot be simulated by the algorithm in Ref. [88]. For these reasons, entanglement was postulated to not be vital for computational speedups achieved by the one clean qubit model [34].

In this work we resolve this question by showing that entanglement is in fact necessary in DQC1. Any circuit from DQC1 that does not generate entanglement can be efficiently classically simulated. We prove this by characterising these circuits. We consider two ways of enforcing that the circuit produce no entanglement; requiring the states to be seperable after each gate applied, or also requiring the state to be seperable throughout the entire computation. Even though circuits

produced in either case are surprisingly nontrivial, we show that a classical simulation of them can be performed.

Our result shows that despite the limited amount of it present in DQC1, entanglement is playing a crucial role in the quantum speedups achieved by the one clean qubit model. This suggests that entanglement may be necessary in other mixed state quantum computers after all.

4.1 The one clean qubit model

The **one clean qubit model** of computation has input state $|+\rangle\langle+|\otimes I/2^n$, where $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Then the clean qubit is used to apply a controlled n qubit unitary U on the mixed register. Note that U must be constructed from a polynomial number of constant size gates[2]. In particular, we will assume that these are 2-qubit gates. Finally, the the originally clean qubit is measured in either the Pauli X or Y basis. It is also possible to define the one clean qubit model so as to allow a circuit to be applied to all the qubits. It can be shown that both definitions give rise to the same complexity class (see e.g. Ref [79] for a constructive proof of this fact) and so we will restrict our attention to the above formulation throughout this work. To be explicit: when we refer to the one clean qubit model, we mean a circuit of the form $|0\rangle\langle 0|\otimes I + |1\rangle\langle 1|\otimes U$ is applied to the state $|+\rangle\langle+|\otimes I/2^n$, and the first qubit is measured in the X or Y basis.
To understand why this model can compute the normalised trace, let us consider the decomposition of the maximally mixed state

$$\frac{I}{2^n} = \sum_i \frac{1}{2^n} |u_i\rangle\langle u_i|, \tag{4.1}$$

where $\{|u_i\rangle\}_i$ is an orthonormal basis formed of eigenvectors of U.
For convenience we will use the notation $\{p(i), |\psi_i\rangle\}_i$, where p is a probability distribution, to denote the mixed state $\sum_i p(i)|\psi_i\rangle\langle\psi_i|$. This notation highlights that the state can be thought of as a probabilistic mixture of pure states, but that the ensemble is generally not unique.
In this case the above equation implies that the initial state can be considered to be $\{\frac{1}{2^n}, |+\rangle|u_i\rangle\}_i$. Let λ_i be the eigenvalue of $|u_i\rangle$. Under the action of controlled U the state goes to

$$|+\rangle|u_i\rangle \to \frac{|0\rangle|u_i\rangle + \lambda_i|1\rangle|u_i\rangle}{\sqrt{2}} = \frac{|0\rangle + \lambda_i|1\rangle}{\sqrt{2}}|u_i\rangle. \tag{4.2}$$

If this (pure) state was measured in the X (or Y) basis the expectation value would be $\frac{1}{2} + \frac{1}{2}\mathrm{Re}(\lambda_i)$ (or $\frac{1}{2} + \frac{1}{2}\mathrm{Im}(\lambda_i)$). Because we must average over all eigenvectors in the basis, the actual quantity that is estimated this way is $\frac{1}{2} + \frac{1}{2}\sum_i \lambda_i/2^n = \frac{1}{2} + \frac{1}{2}\mathrm{Tr}(U)/2^n$. The quantity $\mathrm{Tr}(U)/2^n$ is called the normalised trace, and the above method estimates it to inverse polynomial additive error.

[2] If $U = U_m...U_1$, then control U can be constructed by applying control U_1, control U_2 etc.

This does not allow us to compute the trace itself very accurately, but nevertheless appears to be a difficult quantity to compute classically.
We can use this model to define the class of decision problems DQC1 – the class of decision problems that can be decided correctly with probability $1/2+\varepsilon$ using the one clean qubit model, where ε is at most inverse polynomially small.
The notion of *efficient classical simulation* (and its shorthand, *classically simulable*) that we use in this work is the following: for some uniform family of quantum circuits $\mathcal{F}_n$ acting on the $n+1$-qubit state $\rho := |+\rangle\langle+| \otimes \frac{I}{2^n}$, we say that the family $\mathcal{F}_n$ can be efficiently classically simulated if, for any circuit from $\mathcal{F}_n$, we can estimate the probability $p_X(1)$ (or $p_Y(1)$) of obtaining outcome 1 when measuring the on the clean qubit in the X (or Y) basis at the end of the circuit up to additive error $1/O(\text{poly}(n))$ in time $O(\text{poly}(n))$.

4.2 The one clean qubit model without entanglement

Definition 4.2.1. ***(Separable mixed state)*** *A mixed state is separable on the partition $A|B$ if it is described by an ensemble in this form: $\{p(i), |\psi^i\rangle_A \otimes |\phi^i\rangle_B\}_i$. In words, this means the mixed state is separable across $A|B$ if it can be written as a probabilistic mixture of pure states separable across that cut.*

The following theorem gives our first constraint on the entanglement (or lack thereof) present in the one clean qubit model.

Theorem 4.2.2. ***(From Ref [74])*** *The one clean qubit computations do not have entanglement between the 'clean' qubit and the 'noisy' register at any point in the computation.*

Proof. The pure state in equation 4.2 has no entanglement across this partition. The final state of a one clean qubit computation is a (uniform) probabilistic mixture of these pure states. Hence the state is separable. □

Despite there being no entanglement across this bipartition, there are correlations between the two registers. Also note that there can still be entanglement across other cuts in the one clean qubit model. Therefore we will define the *one clean qubit model without entanglement* as being the one clean qubit model (in the sense described in Section 4.1) with no entanglement across *any* cut during the computation. We will define $\mathsf{DQC1}_{\mathsf{sep}}$ to be the class of decision problems that can be efficiently decided in this model. Note that the separability condition can be enforced two ways. If the circuit is composed of local gates, separability can be required after each gate is applied. Alternatively, if the gates are applied in a continuous manner, say by applying a Hamiltonian evolution, it is natural to enforce separability at all points in time. In this work we consider both possibilities. Theorem 4.2.5 refers to the discrete gate case and Theorem 4.2.6 to the continuous case.

The final state of a one clean qubit model computation (in which a unitary U is applied to the mixed register, controlled on the clean qubit) is separable if and only if the unitary U satisfies a particular condition; namely that it must have a separable eigendecomposition:

Definition 4.2.3. *(U **has a separable eigendecomposition**) We say that an n qubit unitary U has a separable eigendecomposition if (a) there is a set of separable eigenvectors $\{|u_j\rangle\}_j$ and (b) there is a probability distribution p such that $\sum_j p_j |u_j\rangle\langle u_j| = I/2^n$.*

This definition implies that these separable eigenvectors $\{|u_j\rangle\}_j$ span the full Hilbert space. However, they need not form a (orthonormal) basis. In general the set is over-complete. However, this condition implies this set of eigenvectors is in some sense spread 'evenly'.

Theorem 4.2.4. *The final state of a one clean qubit model computation (in which the circuit applied to the qubits is of the form $|0\rangle\langle 0| \otimes I + |0\rangle\langle 0| \otimes U$) has no entanglement if and only if U has a separable eigendecomposition.*

Proof. We prove this theorem in Section 4.5. □

When considering a circuit in the one clean qubit model without entanglement, we need to decompose the $n+1$ qubit unitary control U into smaller gates. Suppose that $U = U_r ... U_1$, where U_1 to U_r are 1 and 2 qubit gates. Then control U is composed of control U_1 to control U_r. In this paper we often describe circuits in the one clean qubit model by referring to these 1 and 2 qubit gates U_1 to U_r, with the understanding that it is the controlled versions of these gates that actually needs to be applied.

Theorem 4.2.5. *Let $\mathcal{C}$ be a circuit in $\mathsf{DQC1}_{\mathsf{sep}}$. In this circuit control U is applied, where $U = U_r ... U_1$. Then the following unitaries must have a separable eigendecomposition: U_1, $U_2U_1, U_3U_2U_1$,..., $U_r ... U_1$.*

Theorem 4.2.6. *Let $\mathcal{C}$ be a circuit in $\mathsf{DQC1}_{\mathsf{sep}}$. Suppose the controlled gate applied after time t is $U(t)$, $U(0) = I$ and at time T the full gate is applied, $U(T) = U$. Then for all $0 \leq t \leq T$, $U(t)$ must have a separable eigendecomposition.*

Proof. The theorems follow from enforcing the separability condition after each gate (or at every point in time in the continuous version), and applying Theorem 4.2.4. □

4.3 $\mathsf{DQC1}_{\mathsf{sep}}$ circuits

Though Theorems 4.2.5 and 4.2.6 characterise the circuits that make up $\mathsf{DQC1}_{\mathsf{sep}}$, this characterisation is not explicit. In this section we demonstrate which 1 and 2 qubit gates can be used to construct the circuits. We will start with an illustrative example of a 2 qubit gate that has a separable eigendecomposition.

Lemma 4.3.1. *Gates in the following form have a separable eigendecomposition:*

$$\begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}, \tag{4.3}$$

where B and C are 1 *qubit unitaries. This gate applies B to the second qubit if the first qubit is in state* $|0\rangle$ *and C if the first qubit's state is* $|1\rangle$*.*

Proof. The separable eigendecomposition of this unitary is $\{|0\rangle|b\rangle, |0\rangle|b^{\perp}\rangle, |1\rangle|c\rangle, |1\rangle|c^{\perp}\rangle\}$, where $|b\rangle$ and $|b^{\perp}\rangle$ are eigenvectors of B and $|c\rangle$ and $|c^{\perp}\rangle$ are eigenvectors of C. □

We will generalise this example so that the unitaries can be controlled by a basis other than the computational basis:

Definition 4.3.2. ***(Basis-controlled unitary)*** *A* 2-qubit basis-controlled unitary *is a unitary* $U^{\mathcal{A}}_{B,C}$ *such that, if* $\mathcal{A}$ *is the basis (for 1 qubit)* $\{|a\rangle, |a^{\perp}\rangle\}$*, then*

$$U^{\mathcal{A}}_{B,C}: \begin{array}{c} |a\rangle|\psi\rangle \mapsto |a\rangle B|\psi\rangle \\ |a^{\perp}\rangle|\psi\rangle \mapsto |a^{\perp}\rangle C|\psi\rangle \end{array}.$$

$U^{\mathcal{A}}_{B,C}$ *has eigenvectors* $|a\rangle|b\rangle, |a\rangle|b^{\perp}\rangle, |a^{\perp}\rangle|c\rangle, |a^{\perp}\rangle|c^{\perp}\rangle$*, where* $\mathcal{B} = \{|b\rangle, |b^{\perp}\rangle\}$ *and* $\mathcal{C} = \{|c\rangle, |c^{\perp}\rangle\}$ *are the eigenbases of B and C, respectively. We will draw such a gate as in Figure 4.1. Since these gates will always be 2-qubit gates, we will write* $U^{B,C}_{\mathcal{A}}$ *to represent a basis-controlled unitary controlled on the second qubit, and acting on the first.*

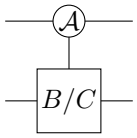


Figure 4.1 A $U^{\mathcal{A}}_{B,C}$ gate.

A continuous time version of this gate can also be constructed:

Observation 4.3.3. ***(Continuous Basis-controlled unitary)*** *Let* $B(t)$ *and* $C(t)$ *be unitary gates such that* $B(0) = C(0) = I$ *and* $B(T) = B$*,* $C(T) = C$*. Then the continuous version of the basis-controlled unitary is* $U^{\mathcal{A}}_{B(t),C(t)}$*, for* $0 \leq t \leq T$*.*

Basis control unitaries are in fact the only 2 qubit unitaries that have a separable eigendecomposition. To formalise this, we prove the following lemma in Section 4.5.

Lemma 4.3.4. *Any 2-qubit unitary that has a separable eigendecomposition must be a basis-controlled unitary $U^{\mathcal{A}}_{B,C}$ (or $U^{B,C}_{\mathcal{A}}$) for some choice of $\mathcal{A}, B$, and C.*

In the next section we will show that $\mathsf{DQC1}_{\mathsf{sep}}$ circuits consist of these 2-qubit basis control unitaries. Before this though, we note two special cases of these gates that will be relevant to us later.

Observation 4.3.5. *Suppose we have the control-unitary $U^{\mathcal{A}}_{B,B'}$, and $[B,B'] = 0$. Then*

- *$U^{\mathcal{A}}_{B,B'}$ is diagonal in the $\mathcal{A}\otimes\mathcal{B}$ basis.*
- *For $B = \begin{pmatrix} e^{i\theta_1} & 0 \\ 0 & e^{i\theta_2} \end{pmatrix}$ and $B' = \begin{pmatrix} e^{i\phi_1} & 0 \\ 0 & e^{i\phi_2} \end{pmatrix}$, we have $U^{\mathcal{A}}_{B,B'} = U^{A,A'}_{\mathcal{B}}$, where $A = \begin{pmatrix} e^{i\theta_1} & 0 \\ 0 & e^{i\phi_1} \end{pmatrix}$ and $A' = \begin{pmatrix} e^{i\theta_2} & 0 \\ 0 & e^{i\phi_2} \end{pmatrix}$.*
- *To emphasis that this gate can be considered to be a control on either line, we will sometimes denote such a gate as $\mathcal{P}_{mathcalAB}$.*

Observation 4.3.6. *Suppose we have the control-unitary $U^{\mathcal{A}}_{B,B'}$, and $B' = e^{i\theta}B$ for some angle θ. Then*

- *$U^{\mathcal{A}}_{B,B'} = U^{\mathcal{A}}_{e^{i\phi}B,e^{i\varphi}B} = A\otimes B$ for $A = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{i\phi} \end{pmatrix}$ in the $\mathcal{A}$ basis. This shows that these 2 qubit gates include all 1 qubit gates as a special case.*

4.3.1 Product control circuits

In this subsection we will define a type of circuit composed of 1 and 2 qubit gates, called a product control circuit, that has a separable eigendecomposition after each gate (and at each point in time in the continuous version). We will also show that this is the only type of circuit with separable eigendecomposition, and hence all circuits in $\mathsf{DQC1}_{\mathsf{sep}}$ are of this form.
Every gate in a product control circuit is a basis-controlled unitary of the type in Definition 4.3.2. However, these gates cannot be placed arbitrarily. Figure 4.2 illustrates an example of a product control circuit.
Informally we can describe the example circuit in Figure 4.2 as follows. The first line is labelled a control line in basis $\mathcal{A}$. As we will see, the only gates that can be applied to a control line is a unitary that uses that line as a control (in basis $\mathcal{A}$). Gates like $\mathcal{P}_{\mathcal{AB}}$ are permitted as they can be considered a control on either line. Target lines in this circuit can only act as targets for basis control unitaries, which we see is the case in this example. Free lines are those that are

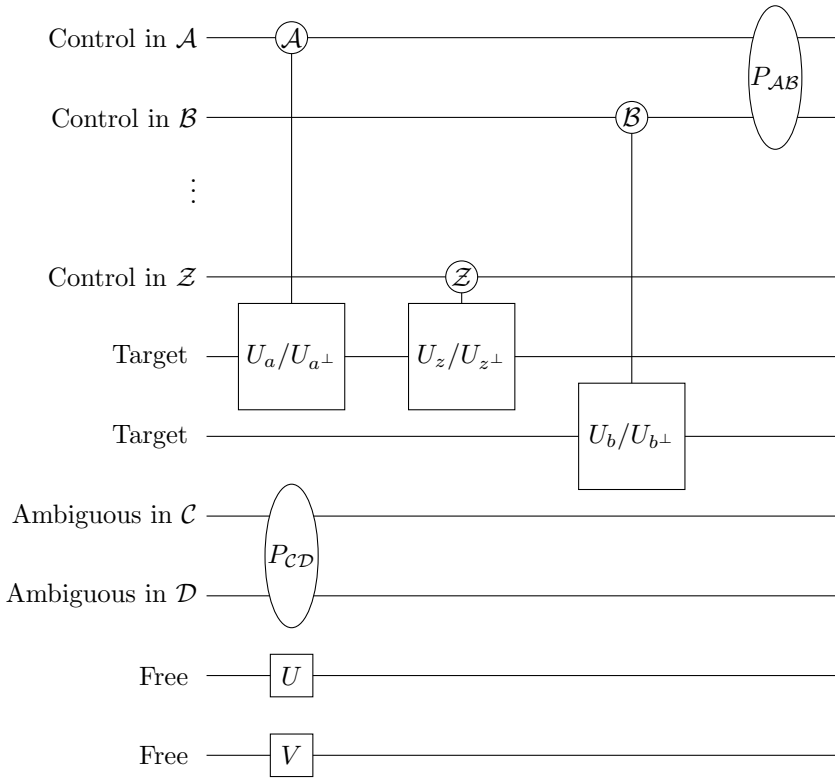


Figure 4.2 This figure gives an example of a circuit that can be constructed using Table 4.1. In this example, the first control line applies control unitaries on target and ambiguous lines using $\mathcal{A}$ as the control basis. Gates that are diagonal in $\mathcal{A} \otimes \mathcal{B}$ can be applied between the first two control lines, and similarly, a gate diagonal in $\mathcal{C} \otimes \mathcal{D}$ is possible between ambiguous lines.

not affected by the states on other lines. In this example, the free lines have only had 1 qubit unitaries applied to them. Ambiguous lines are those that have only had diagonal gates applied to them. They are called ambiguous because a new gate applied in this circuit can treat such a line as either a target or control.
All lines begin as free lines, but may change their classification as new gates are applied. As a trivial example, consider the case where $U^{\mathcal{A}}_{B,C}$, for $[A,B] \neq 0$ is applied to two free lines, making them them a control and target line respectively. However, if $U^{\mathcal{A}}_{B^\dagger,C^\dagger}$ is applied then these lines become free again.

Definition 4.3.7. *All lines begin as free lines. When a new unitary is applied according to what is allowed by Table 4.1, some lines may update their categorisation. This can be determined by the following definitions.*

- *Fixed basis line $\mathcal{A}$: Line i has a fixed basis $\mathcal{A} = \{|0\rangle, |1\rangle\}$ after the unitary is applied if either of the these basis states are inputted on line i (and an arbitrary state is inputted on the other lines), the output on line i is the same basis state.*

- *Control in $\mathcal{A}$: A fixed basis line in $\mathcal{A}$ with the property that a target line k for this line exists, defined as follows. Suppose we are checking if line i is a control line, with target k, where k was not a control line in the previous iteration. Set the input of all previously classified control lines (besides i) to one of their basis states. Call this state $|x\rangle_C$, where x is a binary string representing which choice was made. Set i to input state $|0\rangle$ or $|1\rangle$ (basis states of $\mathcal{A}$). Consider the 1 qubit unitary V_{x0} or V_{x1} applied to line k given that the other lines have their inputs set. If V_{x0} and V_{x1} are not diagonal in a basis $\mathcal{B}$ for all x, then k is a target line of i.*

- *Ambiguous in $\mathcal{B}$: Same condition as for a control, except that V_{x0} and V_{x1} are diagonal in a basis $\mathcal{B}$, and there exists x such that $V_{x0} \neq V_{x1}$.*

- *Target : A line that is targeted by a control line in the sense described above.*

- *Free : If any state $|\psi\rangle$ is inputted on this line and an arbitrary state is inputted on the other lines, the output state on this line is not entangled with the other lines and is independent of the state inputted on those lines.*

The following table explains what unitaries are allowed to be applied between various lines i and j. Generally they are as excepted; control lines can only be used as controls, targets as targets, etc. However, in some cases it is possible to apply a unitary between i and j that will first change the identity of these lines. In our trivial example above, applying $U^{\mathcal{A}}_{B^\dagger,C^\dagger}$ to i and j made them go from a control and target line respectively to free lines. The special cased listed in the table are of this type.

i	j	U	Proof	i after	j after
Control, $\mathcal{A}$	Control, $\mathcal{B}$	diagonal in $\mathcal{A}\otimes\mathcal{B}$	Theorem 4.4.3	Control, $\mathcal{A}$	Control, $\mathcal{B}$
Control, $\mathcal{A}$	Ambiguous, $\mathcal{B}$	$U^{\mathcal{A}}_{K,H}$	Theorem 4.4.3	Control, $\mathcal{A}$	Ambiguous/ target
Control, $\mathcal{A}$	Free	$U^{\mathcal{A}}_{K,H}$	Theorem 4.4.3	Control, $\mathcal{A}$	Free/ ambiguous/ target
Control, $\mathcal{A}$	Target (if special case in Theorem 4.4.3 doesn't hold)	$U^{\mathcal{A}}_{K,H}$	Theorem 4.4.3	Control, $\mathcal{A}$	Target
Control, $\mathcal{A}$	Target (if special case in Theorem 4.4.2 holds)	Either $U^{\mathcal{A}}_{K,H}$, or $U^{K,H}_{\mathcal{B}}U^{\mathcal{A}}_{E,F}$	Theorem 4.4.3	Either control, $\mathcal{A}$, or Target/ Ambiguous	Target Control, $\mathcal{B}$
Target	Target (if cond (i) in Theorem 4.4.5 doesn't hold)	None	Theorem 4.4.5	Target	Target
Target	Target (if cond (i) in Theorem 4.4.5 holds)	Either none, or $U^{\mathcal{A}}_{K,H}(W_x^{\dagger}\otimes I)$	Theorem 4.4.5	Either target Or control, $\mathcal{A}$	Target Target
Target	Free (if cond (ii) in Theorem 4.4.5 doesn't hold)	$U^{H,K}_{\mathcal{A}}$	Theorem 4.4.5	Target	Control, $\mathcal{A}$/ free

i	j	U	Proof	i after	j after
Target	Ambiguous, $\mathcal{B}$ (if cond (iii) in Theorem 4.4.5 doesn't hold)	$U_{\mathcal{B}}^{K,H}$	Theorem 4.4.5	Target/ Ambiguous	Control/ Ambiguous, $\mathcal{B}$
Target	Ambiguous, $\mathcal{B}$ (if cond (iii) in Theorem 4.4.5 holds)	Either $U_{\mathcal{B}}^{K,H}$,	Theorem 4.4.5	Either target/ ambiguous	Control/ Ambiguous, $\mathcal{B}$
		Or $U_{K,H}^{\mathcal{A}}(W_x^\dagger \otimes I)$	Theorem 4.4.5	Or control, $\mathcal{A}$	Target/ Ambiguous, $\mathcal{B}$
Ambiguous, $\mathcal{A}$	Ambiguous, $\mathcal{B}$	Either $U_{K,H}^{\mathcal{A}}$	Theorem 4.4.4	Either ambiguous/ control, $\mathcal{A}$	Ambiguous, $\mathcal{B}$/ target
		Either $U_{\mathcal{B}}^{K,H}$		Or ambiguous, $\mathcal{A}$/ target	Ambiguous/ control, $\mathcal{B}$
Ambiguous, $\mathcal{A}$	Free	Either $U_{K,H}^{\mathcal{A}}$	Theorem 4.4.4	Control/ Ambiguous, $\mathcal{A}$	Target/ Ambiguous
		Or $U_{\mathcal{B}}^{K,H}$		Target/ Ambiguous	Control, $\mathcal{B}$
Free, W	Free, V	$U_{K,H}^{\mathcal{A}}(W^\dagger \otimes V^\dagger)$	Theorem 4.4.5	Control/ Ambiguous, $\mathcal{A}$	Target/ Ambiguous

Table 4.1 This table shows what unitaries are allowed between qubit i and j, depending on their current classification in the circuit. In this table 'Control, $\mathcal{A}$' and 'ambiguous, $\mathcal{A}$' mean a control/ ambiguous line with basis $\mathcal{A}$. 'Free, V' means a free line which has had the 1 qubit unitary V applied to it. This table also shows how the line will be classified after the unitary is applied. However this usually depends on the free parameters in the unitary as well as the previous circuit. These details can be found in the relevant proofs.

Definition 4.3.8. ***(Product control unitary)*** *A n qubit unitary is a product control unitary if it is constructed as a series of* 2 *qubit control unitaries in the following way. All qubits are originally classed as 'free'. A basis-controlled unitary is applied between i and j according to Table 4.1, and the qubits are reclassified accordingly. The process is continued in this way until the full unitary is constructed. A continuous version of this class can be constructed via Observation 4.3.3.*

In the next section we will prove our main technical result, that all $\mathsf{DQC1}_{\mathsf{sep}}$ circuits are product control unitary circuits.

Theorem 4.3.9. *Any* $n+1$*-qubit* $\mathsf{DQC1}_{\mathsf{sep}}$ *circuit in a uniformly family of circuits acts on state* $|+\rangle\langle+| \otimes \frac{1}{2^n}$ *and is necessarily of the form* $|0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes V$ *for some n-qubit product control unitary* V*.*

Definition 4.3.10. ***(Canonical form)*** *An n-qubit circuit is in canonical form if the following conditions are met.*

- *There are some lines that are fixed basis lines. Let us assume that they are lines* 1 *to s and their fixed orthonormal bases are* $\mathcal{B}_i = \{|b_0^{(i)}\rangle, |b_1^{(i)}\rangle\}$*. Collectively we will refer to these lines as B. The other r lines will be referred to as* $T_1, ..., T_r$*.*
- *The circuit is an n qubit unitary* $\mathcal{C} = \sum_x e^{i\theta_x} |x\rangle\langle x|_B \otimes W_{T_1}^{1x} \otimes ... \otimes W_{T_r}^{rx}$*, where* $|x\rangle = |b_{x_1}^{(1)}\rangle \otimes ... \otimes |b_{x_s}^{(s)}\rangle$*. We will refer to this unitary as an n qubit basis control unitary. This circuit can be decomposed as a product control unitary.*

Theorem 4.3.11. *Any uniform family of product control unitaries can be efficiently transformed into a uniform family of canonical form circuits. Furthermore, the fixed basis of each of the lines of the canonical form circuit can also be efficiently computed.*

The proof of this theorem is at the end of Section 4.4.

Theorem 4.3.12. *Any uniform circuit family from* $\mathsf{DQC1}_{\mathsf{sep}}$ *can be simulated efficiently classically. That is,* $\mathsf{DQC1}_{\mathsf{sep}} \subseteq BPP$*.*

Proof. Combining Theorem 4.3.9 and Theorem 4.3.11 we get that any $n+1$-qubit $\mathsf{DQC1}_{\mathsf{sep}}$ circuit in a uniformly family of circuits can be efficiently transformed to a circuit of the form $|0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \mathcal{C}$, where $\mathcal{C}$ is a canonical form circuit as per Definition 4.3.10.
To simulate this circuit it suffices to (uniformly) sample from the eigenvalues of $\mathcal{C}$. In order to do, we first uniformly sample an s-bit string x so that $|x\rangle = |b_{x_1}^{(1)}\rangle \otimes ... \otimes |b_{x_r}^{(s)}\rangle$ is the part of an eigenvector on the fixed basis lines B. The remaining r qubits T are acted on by $W_{T_1}^{1x} \otimes ... \otimes W_{T_r}^{rx}$. For each of these 1-qubit unitaries W^{jx}, compute (in polynomial-time) an eigenbasis $\{|w_0^{(ix)}\rangle, |w_1^{(ix)}\rangle\}$, with corresponding eigenvalues $w_0^{(ix)}, w_1^{(ix)}$. Uniformly sample an r-bit string y and let $|y\rangle = |w_{y_1}^{(1)}\rangle \otimes ... \otimes |w_{y_r}^{(r)}\rangle$.

The vector $|x\rangle_B \otimes |y\rangle_T$ is an eigenvector of $\mathcal{C}$. Moreover, vectors in this form are collectively an orthonormal basis. The corresponding eigenvalues are easy to compute $\mathcal{C}(|x\rangle_B \otimes |y\rangle_T) = e^{i\theta_x} w_{y_1}^{(1x)} ... w_{y_r}^{(rx)} |x\rangle_B \otimes |y\rangle$. Repeating this a polynomial number of times allows us to sample uniformly from the eigenvalues $e^{i\theta_x} w_{y_1}^{(1x)} ... w_{y_r}^{(rx)}$, which in turn allows us to estimate the acceptance probability of the DQC1 circuit up to inverse polynomial additive accuracy.

□

4.4 Main proofs

In this section we prove the results referenced in Theorem 4.3.9. In the proofs that follow, a simple observation about entangled states, which we illustrate with the following example, will prove useful to us many times:

Observation 4.4.1. *Suppose that a state $a|0\rangle_A|\phi_0\rangle_B + b|1\rangle_A|\phi_1\rangle_B$ is not entangled across the bipartition of A and B. Because $|0\rangle_A$ and $|1\rangle_A$ are orthogonal, it must be the case that $|\phi_0\rangle_B$ is proportional to $|\phi_1\rangle_B$. Otherwise it would not be possible to factorise the two components of the vector.*

We state the generalisation of this example rigorously as Lemma 4.6.1 in Section 4.6.
Now we prove our main technical result (Theorem 4.3.9)

Proof. (**Theorem 4.3.9**)
This proof follows by induction. We will assume a product control circuit C has been applied already to the n qubits. Now a new 2-qubit gate U will be applied between i and j. We will show that the only way for the resulting circuit to have a separable eigendecomposition is if U follows the rules in Table 4.1, and hence the resulting circuit must remain of product control form, with the types (i.e. free, control, etc.) of the lines being updated according to the rules of Table 4.1. The conditions that U must obey, and how the types of lines i and j get updated, are given in the rows of the table corresponding to the types of lines i and j.

□

We begin by showing, in Theorem 4.4.2, that when adding a new 2-qubit gate to a circuit C, any (separable) eigenvector of C remains unchanged on those qubits that are not acted upon by that 2-qubit gate. Following this, in Theorem 4.4.3-4.4.5, we consider what constraints this new 2-qubit gate must satisfy in order to preserve the separable eigenbasis condition, and then how the eigenvector changes with respect to the two qubits upon which this gate acts.

Theorem 4.4.2. *Suppose that we have a circuit C on n qubits in the product control form (see Figure 4.2). Now suppose that a 2-qubit gate U is applied to qubits i and j, and that the unitary of the resulting circuit has a separable eigendecomposition $\mathcal{P}$.*

Then for all qubits $k \neq i, j$ corresponding to control, ambiguous, or free lines in basis $\mathcal{B}_k$ (as defined in Definition 4.3.7), the kth qubit of every eigenvector in $\mathcal{P}$ is in the state $|b_0^{(k)}\rangle$ or $|b_1^{(k)}\rangle$ from $\mathcal{B}_k$. (Note that a line could change its status (e.g. from control to ambiguous), but its basis remains unchanged).

Proof. Let S be the set of all control, ambiguous, and free lines other than i and j, and let $S' = [n] \setminus S$ be the rest. Let x be an $|S|$-bit string, and suppose that all qubits in S are in the state $|x\rangle_S := \bigotimes_{k \in S} |b_{x_k}^{(k)}\rangle$. For some state $|\psi\rangle$ on the qubits in S', we have $|\psi\rangle_{S'} \otimes |x\rangle_S \mapsto^C |\psi'\rangle_{S'} \otimes |x\rangle_S$, and we can write the action of C as

$$C = \sum_{x \in \{0,1\}^{|S|}} (C_x)_{S'} \otimes |x\rangle \langle x|_S ,$$

where $(C_x)_{S'}$ is a unitary acting on the qubits in S'. Let $\mathcal{UC}_x$ be an eigenbasis for UC_x. Then $\bigcup_{x \in \{0,1\}^{|S|}} (\mathcal{UC}_x \otimes |x\rangle \langle x|_S)$ is an eigenbasis of UC, and is product if the basis $\mathcal{UC}_x$ is product for all $x \in \{0,1\}^{|S|}$.

Suppose by way of contradiction that there is no such separable eigendecomposition for $UC_{x'}$, for some x', but that there *is* a separable eigendecomposition for the whole circuit UC. Choose an eigenvector $|\phi\rangle_{S'} \otimes |\psi\rangle_S$ from that basis, and write $|\psi\rangle_S = \sum_{x \in \{0,1\}^{|S|}} \alpha_x |x\rangle_S$. Note that there must exist some choice of $|\phi\rangle$ such that $\alpha_{x'} \neq 0$, else the set of eigenvectors wouldn't span the entire space over the n qubits. Without loss of generality, assume that we have chosen such a $|\phi\rangle$.
The circuit UC acts on this eigenvector as

$$|\phi\rangle_{S'} \otimes |\psi\rangle_S \mapsto \sum_{x \in \{0,1\}^{|S|}} \alpha_x (UC_x |\phi\rangle) \otimes |x\rangle_S \quad = \quad e^{i\theta} |\phi\rangle_{S'} \otimes |\psi\rangle_S$$

for some angle θ. By Lemma 4.6.1, the equality can only be true if $|\phi\rangle$ is an eigenvector of all UC_x such that $\alpha_x \neq 0$ (which includes x' by assumption). Since this holds for any eigenvector that we choose from the eigenbasis of UC, this contradicts our assumption that there exists no separable eigendecomposition for UC_x. Hence there must exist a separable eigendecomposition for UC whose eigenvectors are of the form $|\phi\rangle_{S'} \otimes |x\rangle_S$ for all $x \in \{0,1\}^{|S|}$.
In particular, this means that for any qubit $k \neq i, j$ corresponding to a control, ambiguous, or free line in basis $\mathcal{B}$, the kth qubit of any eigenvector of the circuit UC remains unchanged from the corresponding eigenvector of the circuit C. This means a fixed basis line remains a fixed basis line if it is not acted upon. □

The next theorem shows what constraints must be obeyed by a new 2-qubit gate if it is to act on a control qubit i in basis $\mathcal{A}$ and another qubit j. As we show, the new unitary can usually only be a control-unitary controlled on qubit i in basis $\mathcal{A}$ and acting on qubit j. However, if a

certain condition is met, then it is possible for the gate to be a control-unitary of the form $U^{\mathcal{E}}_{B,B'}$ for some basis $\mathcal{E}$ and commuting unitaries B and B'. In this case, by Lemma 4.4.2, the gate can be controlled on either one of i, j, whilst acting on the other. The intuition behind this special case is that, sometimes, the new gate might act to 'cancel' (or more precisely, diagonalise) what has acted on line j so far, which then frees line j up to potentially act as a control on another line, and simultaneously frees line i from its control status.

Theorem 4.4.3. *Suppose that we have a circuit C on n qubits in product control form, and that a 2-qubit gate U is applied to qubits i and j, and also that the unitary of the resulting circuit has a separable eigendecomposition.*
Suppose that qubit i is a control qubit in basis $\mathcal{A}$ (which for simplicity, and wlog, we will assume is just the computational basis). Then, either

(i) There exists a separable eigendecomposition of UC in which line i always has state $|0\rangle_i$ or $|1\rangle_i$, or

(ii) Line j is the only target line that i acts on, and the following condition holds. Let all other control lines be in the state $|x\rangle$ and suppose that C is such that $V_{0,x}$ is applied to j when i is in state $|0\rangle_i$, and $V_{1,x}$ is applied to j when i is in state $|1\rangle_i$. Then there exist unitaries G and H such that both $GV_{0,x}$ and $HV_{1,x}$ are diagonal in some basis $\mathcal{O}$ for all x. In this case, it is possible that the state on line i in the eigendecomposition of UC is an arbitrary single qubit state.

In each case, the new 2-qubit gate is either a basis-controlled unitary itself, or can be written as a sequence of 2 basis-controlled unitaries that satisfy the constraints in Table **??**.
An immediate consequence of these constraints is that the circuit remains in product control form after the addition of the new 2-qubit gate.

Proof. Similarly to the proof of Lemma 4.4.2, we can consider the actions and eigenbases of the circuits UC_x for $x \in \{0,1\}^{|S|}$, where S is the set of all control, ambiguous, and free lines other than i and j. Let T be the set of all target qubits outside of i and j. We divide the proof into two main sections: we start by characterising the form of the eigenvectors and the allowed form of the new 2-qubit gate U in the case that i has previously acted as a control on lines other than j (part 1), and then consider the case in which it acted only on j (part 2).

Part 1: i has previously acted on lines other than j

First, we show that if there is any line other than j that i acts as a control on, then the eigenvectors of the new circuit must have state $|0\rangle_i$ or $|1\rangle_i$ on line i, implying that the new 2 qubit gate must be a control-unitary controlled in basis $\mathcal{A}$, and so the circuit can be re-written so that only a single control-unitary gate is applied between i and j.

Case. Previous circuit on i and j: $C_x = U^{\mathcal{A}}_{V_{0,x},V_{1,x}}$	Form of U	Line i	Line j
(i) Condition (ii) not met.	$U = U^{\mathcal{A}}_{B,C}$ for $BV_{0,x}, CV_{1,x}$ not simultaneously diagonal for all x.	Control in $\mathcal{A}$	Target
	$U = U^{\mathcal{A}}_{E,E'} U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}$ if $V_{0,x}K_0^\dagger$ and $V_{1,x}K_1^\dagger$ are diagonal in some basis $\mathcal{E}$ for all x and E,E' simultaneously diagonal in $\mathcal{E}$.	Control in $\mathcal{A}$	Ambiguous or control in $\mathcal{E}$
(ii) Line i has only acted previously as a control on line j, and it is possible to write $U^{\mathcal{A}}_{V_{0,x},V_{1,x}} = U^{\mathcal{A}}_{K_0,K_1} U^{\mathcal{A}}_{P_{0,x},P_{1,x}}$ where $P_{0,x}$ and $P_{1,x}$ are both diagonal in a basis $\mathcal{E}$ for all x.	$U = U^{\mathcal{A}}_{B,C} U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}$, B,C not simultaneously diagonal in $\mathcal{E}$	Control in $\mathcal{A}$	Target
	$U = U^{\mathcal{A}}_{E,E'} U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}$, E,E' simultaneously diagonal in $\mathcal{E}$	Ambiguous in $\mathcal{A}$	Ambiguous in $\mathcal{E}$
	$U = U^{\mathcal{A},\mathcal{A}'}_{\mathcal{E}} U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}$, $\mathcal{A},\mathcal{A}'$ simultaneously diagonal in $\mathcal{A}$		
	$U = U^{B,C}_{\mathcal{E}} U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}$, B,C not simultaneously diagonal in $\mathcal{A}$	Target	Control in $\mathcal{E}$

Table 4.2 Constraints on the new 2-qubit unitary, depending on the properties of the circuit applied so far, and what happens to the status of the lines i and j after this new gate is added.

Suppose, by way of contradiction, that there is at least one line l in T that i acts as a control on, and that the eigenbasis of the circuit consists of a state $\alpha|0\rangle_i + \beta|1\rangle_i$ on line i. Let $W^{(t)}_{0,x}$ (resp. $W^t_{1,x}$) be the action of C_x on qubit $t \in T$ when qubit i is in state $|0\rangle_i$ (resp. $|1\rangle_i$). For an arbitrary (product) eigenvector $|\psi\rangle_i|\phi\rangle_j|x\rangle_S|\varphi\rangle_T$, then writing $|\psi\rangle_i = \alpha|0\rangle_i + \beta|1\rangle_i$, the action of UC_x on this state is

$$\begin{aligned}(\alpha|0\rangle_i + \beta|1\rangle_i)|\phi\rangle_j|x\rangle_S|\varphi\rangle_T \mapsto \quad & \alpha U(|0\rangle_i V_{0,x}|\phi\rangle_j)|x\rangle_S W^{(1)}_{0,x} \otimes \cdots \otimes W^{(|T|)}_{0,x}|\varphi\rangle_T + \\ & \beta U(|1\rangle_i V_{1,x}|\phi\rangle_j)|x\rangle_S W^{(1)}_{1,x} \otimes \cdots \otimes W^{(|T|)}_{1,x}|\varphi\rangle_T,\end{aligned}$$

Therefore, given that $|\psi\rangle_i|\phi\rangle_j|x\rangle_S|\varphi\rangle_T$ is an eigenvector of UC_c, then we must have

$$\begin{aligned}\alpha U(|0\rangle_i V_{0,x}|\phi\rangle_j) W^{(1)}_{0,x} \otimes \cdots \otimes W^{(|T|)}_{0,x}|\varphi\rangle_T \quad & + \\ \beta U(|1\rangle_i V_{1,x}|\phi\rangle_j) W^{(1)}_{1,x} \otimes \cdots \otimes W^{(|T|)}_{1,x}|\varphi\rangle_T \quad & = e^{i\theta}|\psi\rangle_i|\phi\rangle_j|\varphi\rangle_T.\end{aligned}$$

for some angle θ. By Lemma 4.6.1, this implies that $|\varphi\rangle_T$ is an eigenvector of both $W^{(1)}_{0,x} \otimes \cdots \otimes W^{(|T|)}_{0,x}$ and $W^{(1)}_{1,x} \otimes \cdots \otimes W^{(|T|)}_{1,x}$ with eigenvalue $e^{i\theta}$. Since, by assumption, $|\varphi\rangle$ is a product state $|\varphi^{(1)}\rangle \otimes |\varphi^{(2)}\rangle \otimes \cdots \otimes |\varphi^{(l)}\rangle \otimes \cdots \otimes |\varphi^{(|T|)}\rangle$, then $|\varphi^{(l)}\rangle$ must be an eigenvector of both $W^{(l)}_{0,x}$ and $W^{(l)}_{1,x}$. This means that qubit i does not act (non-trivially) as a control qubit on qubit l, contradicting our assumption above. Hence, one of α or β must be 0, which means that all eigenvectors must be of the form $|0\rangle_i \otimes \cdots$ or $|1\rangle_i \otimes \cdots$.

In this case, the new gate U has preserved the basis on line i, which implies that it must be a basis-controlled unitary of the form $U^{\mathcal{A}}_{B,C}$ for some single-qubit unitaries B,C. To see this, consider the eigenbasis of UC_x for some x. We know that in the case we are considering, any eigenvector from this basis has either $|0\rangle_i$ or $|1\rangle_i$ on the ith line. Without loss of generality,

choose an eigenvector with a $|0\rangle$ on this qubit: $|0\rangle_i |\psi\rangle_j |x\rangle_S |\phi\rangle_T$. The action of the circuit on this state is

$$\begin{aligned} |0\rangle_i |\psi\rangle_j |x\rangle_S |\phi\rangle_T &\mapsto |0\rangle_i V_{0,x} |\psi\rangle_j |x\rangle_S W_{0,x} |\phi\rangle_T \\ &\mapsto U(|0\rangle_i V_{0,x} |\psi\rangle_j) |x\rangle_S W_{0,x} |\phi\rangle_T \\ &= e^{i\varphi} |0\rangle_i |\psi\rangle_j |x\rangle_S |\phi\rangle_T \end{aligned}$$

for some angle φ. This implies that $|\phi\rangle_T$ is an eigenvector of $W_{0,x}$. Moreover, $|0\rangle_i |\psi^\perp\rangle_j |x\rangle_S |\phi\rangle_T$ must also be an eigenvector of UC_x, as must $|1\rangle_i |\theta\rangle_j |x\rangle_S |\phi\rangle_T$ and $|1\rangle_i |\theta^\perp\rangle_j |x\rangle_S |\psi\rangle_T$, for some $|\theta\rangle$. This implies that $U = |0\rangle\langle 0|_i \otimes B + |1\rangle\langle 1|_i \otimes C = U^{\mathcal{A}}_{B,C}$, with $|\psi\rangle, |\psi^\perp\rangle$ eigenvectors of $BV_{0,x}$ and $|\theta\rangle, |\theta^\perp\rangle$ eigenvectors of $CV_{1,x}$.

Part 2: i has only ever acted on line j

Having characterised the form of the eigenvectors in the case that i acts as a control on any line besides j, we will now deal with the remaining cases and assume that the only qubit that i acts as a control on is j. We will show that it is possible for the new 2-qubit gate to (in some sense) 'undo' the gates that have been applied between i and j so far, before acting on them with a basis-controlled unitary. That is, it is possible to write the new gate as the product of 2 basis-controlled unitaries, such that the first of these diagonalises the action of the circuit so far on line j, and the second acts as a new basis-controlled unitary between i and j. In this way, it is possible for either line to change its status.

As we observed before, we can write the action of the circuit so far on qubits i and j as a basis-controlled unitary of the form $U^{\mathcal{A}}_{V_{0,x},V_{1,x}}$ when the other qubits are in the state x. If it is possible to write this as $U^{\mathcal{A}}_{V_{0,x},V_{1,x}} = U^{\mathcal{A}}_{K_0,K_1} U^{\mathcal{A}}_{P_{0,x},P_{1,x}}$, where K_0, K_1 are single-qubit unitaries independent of x, and $P_{0,x}, P_{1,x}$ are single-qubit unitaries that depend on x, but are both simultaneously diagonal in some basis $\mathcal{E}$ for all x, then it is possible that the new 2-qubit gate U can 'undo' what has been applied so far, and change the basis and/or status of either line i or j. More precisely, if this is the case, then we can write the overall circuit UC_x as

$$UC_x = UU^{\mathcal{A}}_{V_{0,x},V_{1,x}} = UU^{\mathcal{A}}_{K_0,K_1} U^{\mathcal{A}}_{P_{0,x},P_{1,x}},$$

which must itself be a basis-controlled unitary, by Lemma 4.5.3 and Theorem 4.4.2. Since we don't know which qubit this unitary will be controlled on, let us simply denote it by G_x, and note that it will be either of the general form $U^{\mathcal{A}}_{B,C}$ or $U^{B,C}_{\mathcal{A}}$. This implies that we can always write U as

$$U = G_x U^{\mathcal{A}}_{P^\dagger_{0,x},P^\dagger_{1,x}} U^{\mathcal{A}}_{K^\dagger_0,K^\dagger_1}.$$

(See Figure 4.3 for clarity). Since this must hold for all x, let us fix one particular x, and write

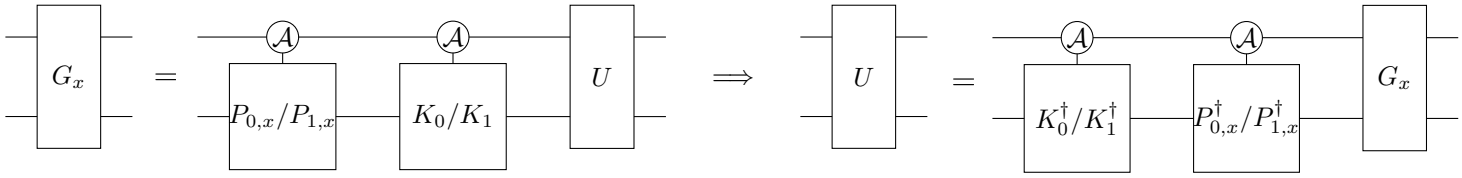


Figure 4.3 $U \cdot U^{\mathcal{A}}_{K_0,K_1} U^{\mathcal{A}}_{P_{0,x},P_{1,x}} = G_x$ implies that $U = G_x U^{\mathcal{A}}_{P_{0,x}^\dagger,P_{1,x}^\dagger} U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}$.

$G = G_x$, $P_0 = P_{0,x}$, and $P_1 = P_{1,x}$. Then we have that

$$UC_x = GU^{\mathcal{A}}_{P_0^\dagger P_{0,x},P_1^\dagger P_{1,x}},$$

and since $P_{0,x}$ and $P_{1,x}$ are diagonal in basis $\mathcal{E}$ for all x, this is simply a gate diagonal in $\mathcal{A}\otimes\mathcal{E}$, followed by a gate G that is a basis-controlled unitary (and indpendent of x), which combine to form the gate G_x. Since we know that G_x is a basis-controlled unitary, it must be the case that G can be combined with $U^{\mathcal{A}}_{P_0^\dagger P_{0,x},P_1^\dagger P_{1,x}}$ to form a basis-controlled unitary. This puts strong constraints on the form that G can take. In particular, since $U^{\mathcal{A}}_{P_0^\dagger P_{0,x},P_1^\dagger P_{1,x}}$ is diagonal in $\mathcal{A}\otimes\mathcal{E}$, then G must act in a way that is consistent with this. This implies that G takes the form $U^{\mathcal{A}}_{B,C}$ or $U^{B,C}_{\mathcal{E}}$ for single-qubit unitaries B,C. The first case is just equivalent to the case discussed in Part 1 of the proof, where line i remains a control line in basis $\mathcal{A}$; however if B and C are diagonal in basis $\mathcal{E}$, then line j becomes an ambiguous line in basis $\mathcal{E}$.

The second case unfolds into similar sub-cases: if $[B,C] \neq 0$, then line i becomes a target line, and line j a control line in basis $\mathcal{E}$; if B and C are diagonal in $\mathcal{A}$, then line i becomes ambiguous in $\mathcal{A}$ and line j ambiguous in $\mathcal{E}$.

Finally, we note that if it is *not* possible to write $U^{\mathcal{A}}_{V_{0,x},V_{1,x}}$ as the product $U^{\mathcal{A}}_{K_0,K_1} U^{\mathcal{A}}_{P_{0,x},P_{1,x}}$, with $P_{0,x},P_{1,x}$ simultaneously diagonal for all x, then the most general form that G_x can take is $U^{\mathcal{A}}_{B,C}$ for non-commuting B,C, which reduces to the case considered in Part 1 of the proof.

All of these cases are considered in Table **??**, where for simplicity we assume that the two gates G and $U^{\mathcal{A}}_{P_0,P_1}$ are combined to form a single basis-controlled unitary. In this table, we also summarise what happens to the status of each line after the new gate is added.

In summary, if i is a control line, then by definition it must have previously acted as a control on at least one other line in the circuit. If it acted on any line other than j, then there is nothing the new gate can do to 'undo' this, and so the new gate can only act as a basis-controlled unitary, controlled on line i and acting on line j. This is shown in part 1 of the proof. In part 2 of the proof, we considered the case that j is the *only* line that i has previously acted on, in which case it is possible for the new gate to act in such a way that it first 'undoes' the action of these previous gates, and then acts with a new basis-controlled unitary. If it fully diagonalises the

Case. Previous circuit on i and j: $C_x = P_{\mathcal{AB}_x} = KQ_{\mathcal{CD}_x}$	Form of U	Line i	Line j
i) $K = I$, $Q_{\mathcal{CD}_x}$ diagonal in $\mathcal{A}\otimes\mathcal{B}$ for all x	$U = U^{\mathcal{A}}_{B,B'}$ for B,B' diagonal in $\mathcal{B}$	Ambiguous in $\mathcal{A}$	Ambiguous in $\mathcal{B}$
	$U = U^{A,A'}_{\mathcal{B}}$ for A,A' diagonal in $\mathcal{A}$		
ii) $K \neq I$, $Q_{\mathcal{CD}_x}$ diagonal in $\mathcal{C}\otimes\mathcal{D}$ for all x	$U = U^{\mathcal{C}}_{E,F}K^\dagger$, $[E,F] \neq 0$	Control in $\mathcal{C}$	Target
	$U = U^{\mathcal{C}}_{D,D'}K^\dagger$, D,D' diagonal in $\mathcal{D}$	Ambiguous in $\mathcal{C}$	Ambiguous in $\mathcal{D}$
	$U = U^{\mathcal{D}}_{C,C'}K^\dagger$, C,C' diagonal in $\mathcal{C}$		
	$U = U^{E,F}_{\mathcal{D}}K^\dagger$, $[E,F] \neq 0$	Target	Control in $\mathcal{D}$

Table 4.3 The possible forms for U, based on the conditions satisfied by the previous circuit, and the resulting status of lines i and j after this new gate is added to the circuit.

previous action of the circuit on line j for all possible states of the other qubits, then it can either act with a full-fledged basis-controlled unitary, controlled on j in that basis, and acting on i as a target; or it can act as a basis-controlled unitary controlled on either of i or j, where the two possible unitaries applied are diagonal in the other line's basis.
In either case it is easy to see that it is possible to write the action of the new gate as a sequence of two basis-controlled unitaries, and so the circuit remains in product-control form.

□

We now deal with the constraints placed on gates that act on ambiguous qubits.

Theorem 4.4.4. *Suppose that the circuit C has already been applied, and now a 2-qubit gate U is applied to qubits i and j, which are both ambiguous in bases $\mathcal{A}$ and $\mathcal{B}$, respectively. Then in order for UC to have a separable eigendecomposition, U must be a basis-controlled unitary, or a sequence of basis controlled unitaries, controlled on either i, in basis $\mathcal{A}$; j, in basis $\mathcal{B}$; or both (see Table 4.3).*

Proof. Once again, we will consider the action of the circuit UC_x for some x. We know that so far all gates will have acted on line i in basis $\mathcal{A}$ and line j in basis $\mathcal{B}$. There may also have been some unitaries acting between them of the form $U^{\mathcal{A}}_{B,B'}$ (for commuting B,B' both diagonal in $\mathcal{B}$), which we can combine into a single gate $P_{\mathcal{AB}_x}$ diagonal in $\mathcal{A}\otimes\mathcal{B}$. Next we note that we can always decompose this gate as the product of two unitaries, one that is independent of x, and one that depends on x, i.e.

$$P_{\mathcal{AB}_x} = KQ_{\mathcal{CD}_x},$$

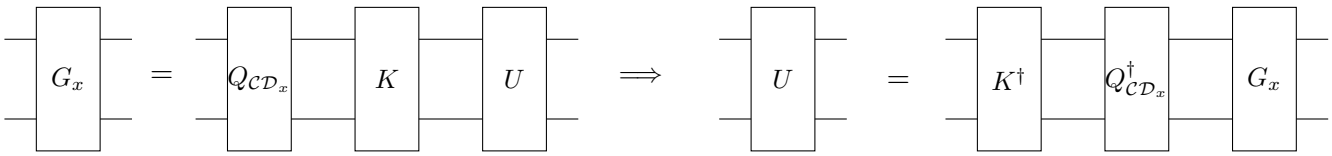


Figure 4.4 $G_x = UKQ_{\mathcal{CD}_x}$ implies that $U = G_x Q^\dagger_{\mathcal{CD}_x} K^d ag$.

where $Q_{\mathcal{CD}_x}$ is diagonal in the basis $\mathcal{C} \otimes \mathcal{D}$ for all x, where it might be the case that $\mathcal{C} = \mathcal{A}$ and/or $\mathcal{D} = \mathcal{B}$, or even that $Q_{\mathcal{CD}_x} = I$ for all x, in which case $\mathcal{C}$ and $\mathcal{D}$ are completely unconstrained.
The resulting circuit UC_x, when restricting attention to qubits i and j, can therefore be written as $UP_{\mathcal{AB}_x} = UKQ_{\mathcal{CD}_x}$. Since this circuit is promised to have a separable eigendecomposition, it must be of control-unitary form. Since we don't yet know which qubit might be the control qubit, we will simply write this basis-controlled unitary as G_x.
These observations imply that we can always write U as

$$U = G_x Q^\dagger_{\mathcal{CD}_x} K^\dagger$$

for any x (see Figure 4.4 for clarity).
Since this holds for all choices of x, we will fix one of these and write $Q_{\mathcal{CD}} := Q_{\mathcal{CD}'_x}$ for some x'. When this is combined with the previous circuit, we obtain the overall unitary $UC_x = GQ^\dagger_{\mathcal{CD}} Q_{\mathcal{CD}_x}$, where the first two gates can be combined to form some unitary $Q'_{\mathcal{CD}_x}$ that is diagonal in the $\mathcal{C} \otimes \mathcal{D}$ basis. We will now consider what form G can take in order to determine what happens to the status of lines i and j, whilst keeping in mind that it must be a basis-controlled unitary. To be a valid basis-controlled unitary when combined with the gate $Q'_{\mathcal{CD}_x}$, then the action on qubit i must be diagonal in $\mathcal{C}$, or the action on qubit j diagonal in $\mathcal{D}$, or both. In the latter case, then G is diagonal in $\mathcal{C} \otimes \mathcal{D}$ and line i remains an ambiguous line, but now in basis $\mathcal{C}$ (which, of course, could be equivalent to $\mathcal{A}$). Likewise, line j remains ambiguous, but now in basis $\mathcal{D}$, with the same observations.
In the first case, we have $G = U^{\mathcal{C}}_{E,F}$ (resp. $G = U^{E,F}_{\mathcal{D}}$) for two single qubit gates E, F. If E and F are both diagonal in the $\mathcal{D}$ (resp. $\mathcal{C}$) basis, then lines i and j both remain ambiguous in bases $\mathcal{C}$ and $\mathcal{D}$, respectively. Otherwise, then for $G = U^{\mathcal{C}}_{E,F}$, line i becomes a control line in $\mathcal{C}$ and line j a target line, and for $G = U^{E,F}_{\mathcal{D}}$, line j becomes a control line in $\mathcal{D}$ and line i a target line.
In summary, we have that U is always of the form $GQ^\dagger_{\mathcal{CD}} K^\dagger$, where K and $Q_{\mathcal{CD}}$ are such that $P_{\mathcal{AB}_x} = KQ_{\mathcal{CD}_x}$ for a fixed choice of x. Then we can combine G and $Q^\dagger_{\mathcal{CD}}$ into a gate G_x that retains the properties of G (i.e. each line will be diagonal (or not) in the same basis), and so the general form of any new 2-qubit gate U added to the circuit must be $U = G'K^\dagger$ (see Table 4.3, and Figure 4.5 for clarity).

□

The remaining cases involve target lines and free lines:

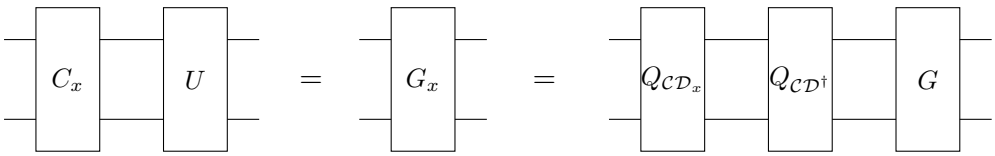


Figure 4.5 $UC_x = G_x = Q_{\mathcal{C}\mathcal{D}_x} Q^\dagger_{\mathcal{C}\mathcal{D}} G$.

Case. Previous circuit on i and j: $C_x = W_x \otimes V_x$	Form of U	Line i	Line j
i) $W_x = KP_x$ for P_x diagonal in basis $\mathcal{C}$ for all x $V_x = LQ_x$ for Q_x diagonal in basis $\mathcal{D}$ for all x Always satisfied when i and j are both target and/or ambiguous lines.	$U = U^{\mathcal{C}}_{E,F}(K^\dagger \otimes L^\dagger)$	Control in $\mathcal{C}$	Target
	$U = U^{\mathcal{C}}_{D,D'}(K^\dagger \otimes L^\dagger)$ for D, D' simultaneously diagonal in $\mathcal{D}$.	Ambiguous in $\mathcal{C}$	Ambiguous in $\mathcal{D}$
	$U = U^{E,F}_{\mathcal{D}}(K^\dagger \otimes L^\dagger)$	Target	Control in $\mathcal{D}$
ii) Only $W_x = KP_x$ for P_x diagonal in basis $\mathcal{C}$ for all x V_x does not have such a decomposition. Always satisfied when one of i and j is a target and/or ambiguous line.	$U = U^{\mathcal{C}}_{E,F}(K^\dagger \otimes V^\dagger)$ for $V = V_x$ for some fixed x.	Control in $\mathcal{C}$	Target
iii) Neither W_x nor V_x admits a decomposition of the form KP_x such that P_x is diagonal for all x.	$U = U_1 \otimes U_2$	Same as before	Same as before

Table 4.4 Possible forms of the new 2-qubit gate U applied to qubits i and j, depending on the constraints satisfied by the circuit applied so far, as well as the new status of the lines after this gate is added.

Theorem 4.4.5. *Applying a 2-qubit gate U to qubits i and j, after applying the product control circuit C, whilst maintaining that the final circuit UC has a separable eigendecomposition places the following constraints on U, and changes the status of lines i and j as shown in Table 4.4. This means the circuit remains in product control form.*

Proof. We will proceed by proving a general statement that applies to all types of non-control lines. I.e. we assume i) that i and j are either ambiguous, target, or free lines, and ii) that no gates have been applied so far between these two lines. If condition i) doesn't hold, then Theorem 4.4.3 applies instead, and if condition ii) doesn't hold, then Theorem 4.4.4 applies.
As usual, we can assume that all other control and ambiguous lines are in some state x, and view the circuit C_x so far as acting with a unitary W_x on qubit i and a unitary V_x on qubit j (in some cases the dependence on x will be redundant – in particular if one of the qubits is a free line). In a similar manner to previous proofs, we make the following set of claims. The overall (2-qubit) unitary acting on qubits i and j after applying the new gate U must have a separable eigendecomposition by Theorem 4.4.2, and therefore be a basis-controlled unitary by Lemma 4.3.4. We don't know which qubit is the control qubit for this basis-controlled unitary, and so we will denote it by G_x. Then we know that $UC_x = G_x = U(W_x \otimes V_x)$, and so we can always write $U = G_x(W_x^\dagger \otimes V_x^\dagger)$. Since this holds for all x, let us fix some arbitrary x and write $U = G(W^\dagger \otimes V^\dagger)$.
Now we consider a number of cases.

Case 1 Suppose that it is possible to to write both $W_x = KP_x$ and $V_x = LQ_x$, where K and L are single qubit unitaries independent of x, and P_x and Q_x are diagonal in some bases $\mathcal{C}$ and $\mathcal{D}$ for all x. Then we can write $U = G(P^\dagger \otimes Q^\dagger)(K^\dagger \otimes L^\dagger)$ (where we once again fixed this decomposition for some arbitrary x), and the action of the entire circuit (on these two qubits) as $UC_x = G_x = G(P^\dagger \otimes Q^\dagger)(K^\dagger \otimes L^\dagger)(K \otimes L)(P_x^\dagger \otimes Q_x^\dagger) = G(P^\dagger \otimes Q^\dagger)(P_x^\dagger \otimes Q_x^\dagger)$. Since P_x and Q_x are diagonal in bases $\mathcal{C}$ and $\mathcal{D}$ for all x, this is the product of a basis-controlled unitary G and two single-qubit, diagonal gates, and we know that this product has to form the basis-controlled unitary G_x. This implies that G must be satisfy one of a number of properties. Either it is of the form $U_{E,F}^{\mathcal{C}}$, $U_{\mathcal{D}}^{E,F}$, or it is diagonal in $\mathcal{C} \otimes \mathcal{D}$. In the first case, if E and F are not simultaneously diagonal in $\mathcal{D}$, then line i becomes a control line in $\mathcal{C}$ and line j a target line. Otherwise, the third case applies. A similar argument applies for the second case. Finally, if G is diagonal in $\mathcal{C} \otimes \mathcal{D}$, then lines i and j become ambiguous in bases $\mathcal{C}$ and $\mathcal{D}$, respectively.

Case 2 Now suppose that it is possible to write only one of W_x or V_x as a decomposition into a part independent of x and a diagonal part dependent on x. Suppose that this is true for W_x, i.e. $W_x = KP_x$, and P_x is diagonal in some basis $\mathcal{C}$ for all x. Now we have fewer options for the form of G. In particular, it can only be of the form $U_{E,F}^{\mathcal{C}}$, for some arbitrary single-qubit unitaries E and F. In this case, qubit i becomes a control line in basis $\mathcal{C}$ and j becomes a target line. A similar argument applies if V_x satisfies the condition instead of W_x.

Case 3 Finally, in the case that neither W_x nor V_x can be decomposed into a part independent of x and a diagonal part dependent on x, then the only form that G can take is as two separate single-qubit unitaries, i.e. $G = U_0 \otimes U_1$.

Now we can consider what conditions can be satisfied when lines i and j are of various types. If i and j are both target lines, then any of the above three cases can apply. If one of them is an ambiguous line, then only cases 1 and 2 can apply. If both are ambiguous, then case 1 always applies. Likewise, if both are free lines, than case 1 always applies. All of this is summarised in Table 4.4, where we have merged the actions of V, W, and G into a single basis-controlled unitary, as appropriate. It is clear that in all cases, the new 2-qubit unitary can be expressed as a combination of at most 2 basis-controlled unitaries, which keeps the circuit in product-control form as required.

□

The above theorems collectively prove that the only circuits with separable eigendecomposition are product control circuit 4.3.9, as shown in Table 4.1. Now we prove Theorem 4.3.11 that shows all product control circuits can be brought into canonical form. A canonical form circuit

contains one n-qubit basis control gate, but this gate can be decomposed into a product control circuit.

Proof. **Theorem 4.3.11:** To prove this theorem we will provide the algorithm for bringing a product control circuit into canonical form efficiently. We will assume that the first d gates in the original circuit have been brought into canonical form. This means they have been combined into one n qubit basis control gate $\mathcal{C} = \sum_x e^{i\theta_x}|x\rangle\langle x|_B \otimes W_{T_1}^{1x} \otimes ... \otimes W_{T_r}^{rx}$. The next gate in the circuit is U and is applied between i and j. We will show that U can be combined with $\mathcal{C}$ to make a new n qubit basis control gate.
The circuit $\mathcal{C}$ is a product control circuit and after applying it the lines in B are fixed basis lines (control/ambiguous/free) and the lines in T are target lines or fixed basis lines for which we could not compute the basis efficiently, but which do not act as controls and so are effectively target lines. To prove this theorem we will consider the different combinations of lines that i and j may be. These cases are the same as those in the proof of Theorem 4.3.9:

1. Line i is a control line (considered in Theorem 4.4.3).
2. Line i is a target line or both i and j are free (considered in Theorem 4.4.5).
3. Line i is ambiguous and j is either ambiguous or free (considered in Theorem 4.4.4).

Case 1: If i is a control line after applying $\mathcal{C}$ then i must be a line in B and its basis is $\mathcal{A}$. If j is not the only table line that i has controlled on then in Table **??** we see that the unitary U must be a basis control unitary controlled on i (in basis $\mathcal{A}$). If j is a control line in basis $\mathcal{B}$ then U is diagonal in $\mathcal{A}\otimes\mathcal{B}$. In this case $U\mathcal{C}$ is clearly still an n qubit basis control unitary. Otherwise if j was ambiguous in basis $\mathcal{B}$ or free then either U is diagonal in $\mathcal{A}\otimes\mathcal{B}$ or j is now a target line. $U\mathcal{C}$ is still an n qubit basis control gate.
Now we consider the first of the special conditions. Suppose j is a target line, $\mathcal{C}_x = U^{\mathcal{A}}_{V_{0,x},V_{1,x}}$, and there exists K_0 and K_1 such that $V_{0,x}K_0^\dagger$ and $V_{1,x}K_1^\dagger$ are diagonal in basis $\mathcal{E}$ for all x. Then U can be of the form $U = U^{\mathcal{A}}_{E,E'}U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}$. Then $U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}\mathcal{C}_x$ is diagonal in basis $\mathcal{A}\otimes\mathcal{E}$. In this case, $U\mathcal{C}$ is a basis control unitary and j is now an ambiguous line with basis $\mathcal{E}$. If this line is ever used as an ambiguous line in the rest of the circuit, i.e. there is unitary controlled on this line later, then it is easy to compute what $\mathcal{E}$ is. Otherwise we will effectively consider j to be a target line.
Suppose j is a i's only target line, $\mathcal{C}_x = U^{\mathcal{A}}_{K_0P_{0,x},K_1P_{1,x}}$, such that $P_{0,x}$ and $P_{1,x}$ are diagonal in basis $\mathcal{E}$ for all x. Then $U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}\mathcal{C}_x$ is diagonal in basis $\mathcal{A}\otimes\mathcal{E}$, and so $U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}\mathcal{E}$ is a basis control unitary in which i and j are ambiguous. $U = U'U^{\mathcal{A}}_{K_0^\dagger,K_1^\dagger}$, where U' either acts as a control on line i or j or both. In any of these cases the resulting unitary is a basis control unitary.
Case 2: If both i and j are free then applying U certainly maintains that the full circuit is canonical. If i is a target line then the circuit so far on these two lines is $C_x = W_x \otimes V_x$. Suppose

$W_x = KP_x$ for P_x diagonal in basis $\mathcal{C}$ for all x. Then applying $K^\dagger$ to that line makes it an ambiguous line. Similarly for line j. Then the cases proceed as above. Otherwise $U = U_1 \otimes U_2$ and so clearly the circuit maintains canonical form.
Case 3: If $K = I$ in the condition for Theorem 4.4.4 then the cases are as above. Otherwise if $K \neq I$, it is possible to apply $K^\dagger$ to make the lines ambiguous in the basis $\mathcal{C} \otimes \mathcal{D}$ instead. Then the cases precede as before.

□

4.5 Gates in $\mathsf{DQC1}_{\mathsf{sep}}$

In this section we discuss properties of gates in $\mathsf{DQC1}_{\mathsf{sep}}$, and we provide the proof of Theorem 4.2.4 and Lemma 4.3.4

Theorem 4.2.4 *The final state of a one clean qubit model computation (in which the circuit applied to the qubits is of the form $|0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes U$) has no entanglement if and only if U has a separable eigendecomposition, as per Definition 4.2.3.*

Proof. The 'if' direction is clear. Equation 4.2 has no entanglement if $\{|u_i\rangle\}_i$ is a separable eigendecomposition. The final state is a mixture of these states and so it does not have entanglement in this case.
For the 'only if' direction, we will first show that for any ensemble representing the initial state, $\{p_i, |+\rangle|\phi_i\rangle\}_i$, the clean and noisy register become entangled unless each vector $|\phi_i\rangle$ is an eigenvector of U [3].
Suppose $\{p_i, |+\rangle|\phi_i\rangle\}_i$ is an ensemble for $|+\rangle\langle+| \otimes I/2^n$. Then the pure states in this ensemble evolve in the following way:

$$|+\rangle|\phi_i\rangle \to \frac{1}{\sqrt{2}}|0\rangle|\phi_i\rangle + \frac{1}{\sqrt{2}}|1\rangle U|\phi_i\rangle. \tag{4.4}$$

Let A be the clean register, and B the noisy one. We wish to show that this state has no entanglement between A and B if and only if $|\phi_i\rangle$ is an eigenvector of U.
As a density matrix, this state is

$$\rho_{AB} \equiv \frac{1}{2}|0\rangle\langle 0| \otimes |\phi_i\rangle\langle\phi_i| + \frac{1}{2}|0\rangle\langle 1| \otimes |\phi_i\rangle\langle\phi_i|U^\dagger + \frac{1}{2}|1\rangle\langle 0| \otimes U|\phi_i\rangle\langle\phi_i| + \frac{1}{2}|1\rangle\langle 1| \otimes U|\phi_i\rangle\langle\phi_i|U^\dagger. \tag{4.5}$$

And the reduced state on A is

$$\rho_A = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|0\rangle\langle 1|\langle\phi_i|U^\dagger|\phi_i\rangle + \frac{1}{2}|1\rangle\langle 0|\langle\phi_i|U|\phi_i\rangle + \frac{1}{2}|1\rangle\langle 1|. \tag{4.6}$$

[3] The ensemble may be continuous here. This will not change the analysis.

This reduced state is pure if and only if ρ_{AB} has no entanglement between A and B. A calculation shows $\mathrm{Tr}(\rho_A^2) = 1/2 + 1/2|\langle\phi_i|U|\phi_i\rangle|^2$, which is 1 if and only if $|\phi_i\rangle$ is an eigenvector of U.
This means that the only ensembles without entanglement between the clean and noisy qubits are ones in which the pure states that make up the ensemble are of the form $|+\rangle|u_j\rangle$, where $|u_j\rangle$ is an eigenvector of U. In particular, this rules out the possibility of the existence of any equivalent ensemble that is *not* of this form, implying that any separable decomposition must consist of pure states $|+\rangle \otimes |b_0\rangle \otimes |b_1\rangle \otimes \cdots \otimes |b_n\rangle$ for single qubit basis states $|b_i\rangle$. Hence, the eigenvector $|u_j\rangle$ themselves must be product across all bipartitions.
Moreover, by applying the inverse of control U on this ensemble, we would recover the maximally mixed state on n qubits, which implies that

$$\sum_j p_j |u_j\rangle\langle u_j| = \frac{I}{2^n}, \tag{4.7}$$

and so U has a separable eigendecomposition as per Definition 4.2.3.
For this to be true $\{|u_j\rangle\}_j$ must necessarily span the full Hilbert space. □

For 2 qubit unitaries, the separable eigendecomposition condition simplifies:

Lemma 4.5.1. *A 2-qubit unitary U has a separable eigendecomposition if and only if there exists a separable orthonormal basis of eigenvectors of U.*

Proof. The eigenvectors in the separable eigendecomposition may not be orthonormal. In the case U has non degenerate eigenvalues then this will not happen. Suppose that U has a degenerate eigenspace of dimension 2. If there are only 2 separable eigenvectors (up to multiplicative factors) in that space then they must be orthonormal. That is because otherwise no mixture of them will equal the identity on that subspace. So suppose there are more than 3 separable eigenvectors in the subspace. Choose 3 of them $|\psi^1\rangle = |\psi_1^1\rangle|\psi_2^1\rangle$, $|\psi^2\rangle = |\psi_1^2\rangle|\psi_2^2\rangle$, $|\psi^3\rangle = |\psi_1^3\rangle|\psi_2^3\rangle$, such that no two vectors are a multiple of each other. Any 2 of these vectors must then space the 2 dimensional subspace.
Let $|\phi\rangle = |\phi_1\rangle|\phi_2\rangle =$ be a separable eigenvector in one of the other eigenspaces. The three above vectors must be orthonormal to $|\phi\rangle$. This is only possible if their state on qubit 1 is $|\phi_1^\perp\rangle$ or their state on qubit 2 is $|\phi_2^\perp\rangle$. Without loss of generality, assume the following:

$$\begin{aligned} |\psi^1\rangle &= |\phi_1^\perp\rangle|\psi_2^1\rangle \\ |\psi^2\rangle &= |\phi_1^\perp\rangle|\psi_2^2\rangle \\ |\psi^3\rangle &= |\psi_1^3\rangle|\phi_2^\perp\rangle \end{aligned} \tag{4.8}$$

The span of $|\psi^1\rangle$ and $|\psi^2\rangle$ therefore contains $|\phi_1^\perp\rangle|\phi_2\rangle$ and $|\phi_1^\perp\rangle|\phi_2^\perp\rangle$. Hence a separable and orthonormal basis exists for this subspace.

If instead U has a degenerate eigenspace of dimension 3, let $|\psi^1\rangle$, $|\psi^1\rangle$, $|\psi^1\rangle$ be 3 separable eigenvectors that span that space, and $|\phi\rangle$ be a separable vector in another subspace. Equation 4.8 still applies, and $|\phi_1^\perp\rangle|\phi_2\rangle$ and $|\phi_1^\perp\rangle|\phi_2^\perp\rangle$ is in the span of these vectors. If these 3 vectors span a 3 dimensional space, it must be the case that $\langle\phi_1|\psi_1^3\rangle \neq 0$. Hence $|\phi_1\rangle|\phi_2^\perp\rangle$ is in the span of these vectors as well. Therefore, an separable orthonormal basis exists.

□

Lemma 4.5.2. *Suppose we have a 2-qubit unitary with a separable eigendecomposition. Then this unitary must be a basis-controlled unitary $U^{\mathcal{A}}_{B,C}$ (or $U^{B,C}_{\mathcal{A}}$) for some choice of $\mathcal{A}, B$, and C.*

Proof. Using Lemma 4.5.1, there is a separable eigenbasis:

$$\{|\psi_0\rangle|\phi_0\rangle, |\psi_1\rangle|\phi_1\rangle, |\psi_2\rangle|\phi_2\rangle, |\psi_3\rangle|\phi_3\rangle\}$$

However, as we will now show, the possible choices for the ψ_i and ϕ_i are heavily constrained. We can pick an eigenbasis by arbitrarily choosing a basis vector $|a\rangle|c\rangle$, and then choose the subsequent basis vectors under the constraint that they must be orthogonal to all previous basis vectors – see Figure 4.6. By inspection, one can see that all choices of bases consistent of a single

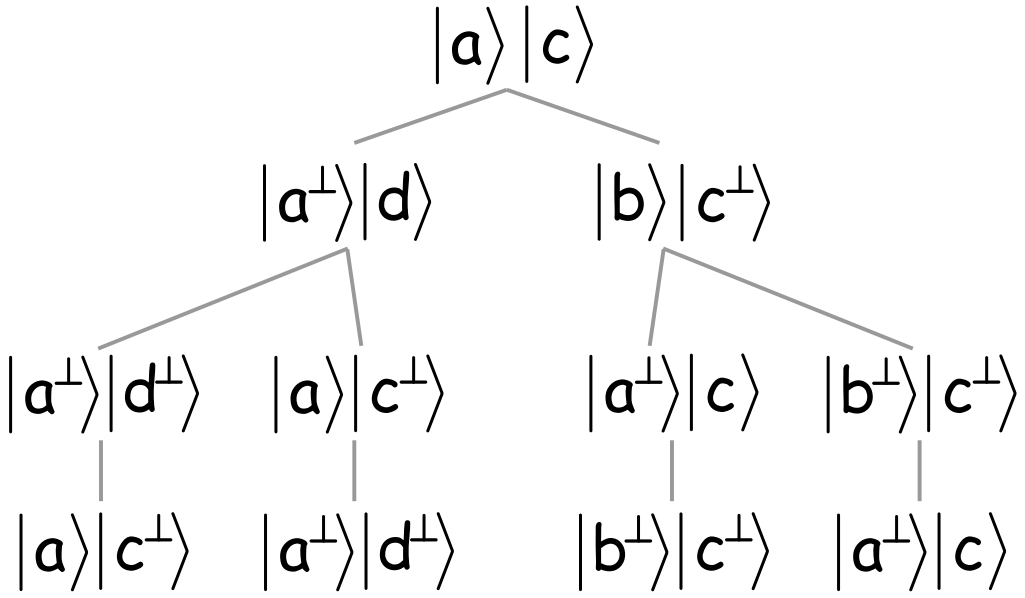


Figure 4.6 All possible choices of product eigenbases for a 2-qubit unitary.

1-qubit basis on one of the qubits, and two 1-qubit bases on the other qubit, each corresponding to one of the possible basis vectors on the first qubit. Hence, the only basis choice we can have is

$$\{|a\rangle|b\rangle, |a\rangle|b^\perp\rangle, |a^\perp\rangle|c\rangle, |a^\perp\rangle|c^\perp\rangle\},$$

for some 1-qubit bases $\{|a\rangle, |a^\perp\rangle\}, \{|b\rangle, |b^\perp\rangle\}, \{|c\rangle, |c^\perp\rangle\}$ (up to swapping the first and second qubit). Hence, the unitary U must be of the form

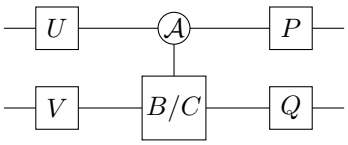
$$|a\rangle\langle a|\left(e^{i\theta_1}|b\rangle\langle b| + e^{i\theta_2}|b^\perp\rangle\langle b^\perp|\right) + |a^\perp\rangle\langle a^\perp|\left(e^{i\phi_1}|c\rangle\langle c| + e^{i\phi_2}|c^\perp\rangle\langle c^\perp|\right),$$

where $e^{i\theta_1}, e^{i\theta_2}, e^{i\phi_1}, e^{i\phi_2}$ are the eigenvalues associated with each eigenvector. Written in the $\mathcal{A} = \{|a\rangle, |a^\perp\rangle\}$ basis, this is

$$\begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix},$$

corresponding to a controlled B/C gate: if qubit 1 is in state $|a\rangle$, then the unitary $B = e^{i\theta_1}|b\rangle\langle b| + e^{i\theta_2}|b^\perp\rangle\langle b^\perp|$ is applied, else if qubit 1 is in state $|a^\perp\rangle$, then unitary $C = e^{i\phi_1}|c\rangle\langle c| + e^{i\phi_2}|c^\perp\rangle\langle c^\perp|$ is applied. Hence, the unitary must be a basis-controlled unitary $U^{\mathcal{A}}_{B,C}$. □

Lemma 4.5.3. *Suppose we have the following circuit:*



Then its corresponding unitary has a separable eigendecomposition in the following cases:

- *If $[QBV, QCV] \neq 0$, and U and P are diagonal in the $\mathcal{A}$ basis.*
- *If $[QBV, QCV] = 0$.*
- *If $B = e^{i\phi}C$ for some angle ϕ.*

Proof. We begin by absorbing V and Q into the basis-controlled unitary. Let $B' := QBV$ and $C' := QCV$. Assume that B' and C' do not commute, and (wlog) that $\mathcal{A}$ is the computational basis (the same result can be obtained by writing a and $a^\perp$ in place of 0 and 1 in what follows). Choose some eigenvector $|\phi_1\rangle|\phi_2\rangle$ of the circuit, and write $U|\phi_1\rangle = c|0\rangle + d|1\rangle$. Then the action of the circuit on this eigenvector is

$$|\phi_1\rangle|\phi_2\rangle = c|0\rangle|\phi_2\rangle + d|1\rangle|\phi_2\rangle \mapsto c|0\rangle B'|\phi_2\rangle + d|1\rangle C'|\phi_2\rangle \mapsto cP|0\rangle B'|\phi_2\rangle + dP|1\rangle C'|\phi_2\rangle.$$

Since P is unitary, $P|0\rangle$ and $P|1\rangle$ are orthogonal, and so the final state is only product if either $c = 0$ or $d = 0$, or if $B'|\phi_2\rangle \propto C'|\phi_2\rangle$. In the latter case, by Lemma 4.6.1, $|\phi_2\rangle$ must be an eigenvector of both B' and C', which implies that $[B', C'] = 0$. By assumption, this is not the case, and so we must have that $c = 0$ or $d = 0$. This implies that $U = P^\dagger$, and also that $|\phi_2\rangle$ is an

eigenvector of at least one of B' and C'.

Now we consider the case where $[B',C'] = 0$. In this case, we know from Observation 4.3.5 that $U^{\mathcal{A}}_{B',C'} = U^{A,A'}_{\mathcal{E}}$, where $\mathcal{E}$ is the shared eigenbasis of B' and C' and A,A' are as defined in Observation 4.3.5. Moreover, we can absorb U and P into the basis-controlled unitary, yielding $U^{PAU,PA'U}_{\mathcal{E}}$.

Finally, if $B = e^{i\phi}C$ for some angle ϕ, then from Observation 4.3.6 we know that the circuit collapses to two sets of 3 unitaries acting on each qubit separately. □

The following corollary follows immediately from Lemma 4.3.4, and demonstrates how to construct a single 2-qubit basis-controlled unitary in each of the three cases above.

Corollary 4.5.4. *If the circuit from Lemma 4.5.3 has a separable eigendecomposition, then it can be written as a single basis-controlled unitary.*

Proof. The correctness follows from Lemma 4.3.4. To find the correct form for the basis-controlled unitary, we observe that if $[QCV, QBC] = 0$, then we can simply absorb the unitaries U,V,P,Q into $U^{\mathcal{A}}_{B,C}$ as in the second half of the proof of Lemma 4.5.3, yielding a basis-controlled unitary $U^{A,A'}_{\mathcal{E}}$. If $B = e^{i\phi}C$ for some angle ϕ, then we can write the entire circuit as two single-qubit unitaries $UAP \otimes QBV$, where $A = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}$. Finally, if $[QCV, QBV] \neq 0$, then U and W must be diagonal in the $\mathcal{A}$ basis, and so they just contribute a relative phase $e^{i\theta}$ which can be absorbed into B and C, yielding the basis-controlled unitary $U^{\mathcal{A}}_{e^{i\theta}QBV, e^{i\theta}QCV}$ □

It is possible to find a series of gates in a product circuit that, when combined act as large controlled unitary. The following Lemma pertains to the case when this gate happens to be diagional in a product basis, as illustrated in Figure 4.7. We can replace the gate with an equivalent gate that is controlled on a different set of qubits, and that acts on a different target qubit, but still with a unitary that is diagonal.

Lemma 4.5.5. *Suppose that we have an n-qubit circuit composed of a single target qubit i, with $n-1$ control unitaries acting on it, each controlled on different control lines (which can be controlled in different bases). Let the basis on the kth line be $\mathcal{B}_k = \{|b^k_0\rangle, |b^k_1\rangle\}$ (where $|b^k_1\rangle = |b^{k\perp}_0\rangle$). Let $x = x_1, x_2, \ldots, x_{n-1}$ be an $(n-1)$-bit string, and $U^{(i)}_x$ be the unitary applied to qubit i when the other control lines are in the state $|x\rangle = |b^1_{x_1}\rangle \otimes |b^2_{x_2}\rangle \otimes \cdots \otimes |b^n_{x_{n-1}}\rangle$.*
Further suppose that $U^{(i)}_x$ is diagonal (in the $\mathcal{B}_i$ basis) for all x. Then the circuit can be re-written as a basis-controlled unitary acting on any qubit $j \in [n]$, controlled on all the others, such that the unitary $U^{(j)}$ acting on the new target qubit j remains diagonal in the basis $\mathcal{B}_j$ of that qubit.

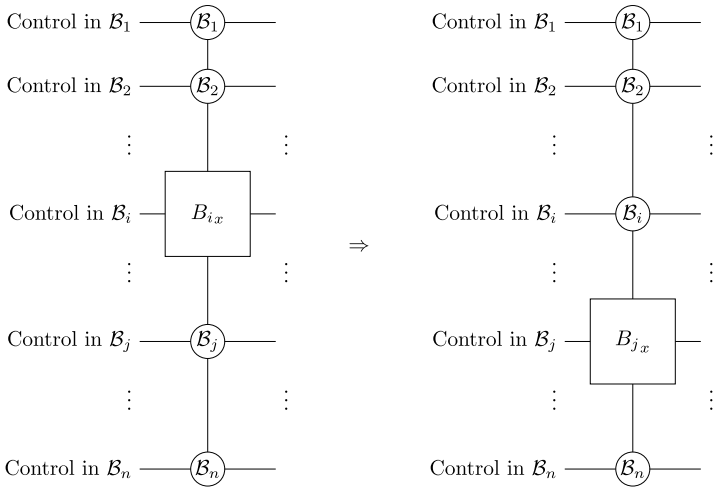


Figure 4.7 Changing the target qubit for a multi-controlled diagonal gate.

Proof. Let the eigenvalues of $|b_0^0\rangle|x\rangle$ and $|b_1^{0\perp}\rangle|x\rangle$ be $e^{i\theta_{0,x}} = e^{i\theta_{1,x_0,x_1,\ldots,x_{n-1}}}$ and $e^{i\theta_{1,x_0,x_1,\ldots,x_{n-1}}}$. Now let $|y,0\rangle$ (resp. $|y,1\rangle$) be the state where all lines but the jth are in state $|y\rangle$, and the jth qubit is in state $|b_0^j\rangle$ (resp $|b_1^j\rangle$). That is, the qubits are in the state $|b_{y_0}^0\rangle \otimes \cdots \otimes |b_b^j\rangle \otimes \cdots \otimes |b_{y_n}^n\rangle$ for $b = 0$ (resp. $b = 1$). Then B_y has eigenvalues $e^{i\theta_{y_1,\ldots,0_j,\ldots,y_n}}$ and $e^{i\theta_{y_1,\ldots,1_j,\ldots,y_n}}$ corresponding to eigenvectors $|y,0\rangle$ and $|y,1\rangle$, respectively.
Hence, B_y is diagonal in basis $\mathcal{B}_j$. Furthermore, given that one knows this condition is met, it is efficient to compute B_y for any y. □

4.6 Formal statement and proof of Observation 4.4.1

Lemma 4.6.1. *Suppose we have a state on $|C|+|T|$ qubits, acted upon by a unitary $U = \sum_x |x\rangle\langle x| \otimes U_x$. Suppose this unitary has a product eigenvector $|\psi\rangle_C|\phi\rangle_T = \sum_x \alpha_x |x\rangle_C|\phi\rangle_T$ with eigenvalue $e^{i\theta}$:*

$$\sum_x \alpha_x |x\rangle_C |\phi\rangle_T \mapsto U\left(\sum_x \alpha_x |x\rangle_C |\phi\rangle_T\right) = \sum_x \alpha_x |x\rangle_C U_x |\phi\rangle_T .$$

Then for all $\alpha_x \neq 0$, $|\phi\rangle_T$ is an eigenvector of U_x with eigenvalue $e^{i\theta}$.

Proof. We have

$$\sum_x \alpha_x |x\rangle_C U_x |\phi\rangle_T = e^{i\theta}\left(\sum_x \alpha_x |x\rangle_C |\phi\rangle_T\right)$$

for some phase θ. Write $|\phi\rangle = \sum_y \beta_y |y\rangle$. Then

$$\sum_{x,y} \alpha_x \beta_y |x\rangle_C U_x |y\rangle_T = e^{i\theta} \sum_{x,y} \alpha_x \beta_y |x\rangle |y\rangle .$$

In particular, this means that for every x, y,

$$|x\rangle U_x |y\rangle = e^{i\theta} |x\rangle |y\rangle$$

Since $\{|x\rangle |y\rangle\}_{x,y}$ forms a orthonormal basis over both registers, this condition can only hold if $U_x |y\rangle = e^{i\theta} |y\rangle$ for all x, y. □

4.7 Remarks and future work

In this work we have shown that the One Clean Qubit model without entanglement is classically simulable. This leaves open the larger question: are all (mixed state) quantum computers classically simulable without entanglement? We conjecture the following.

Conjecture 4.7.1. *Every uniformly constructed family of circuits without entanglement can be classically simulated. That is, for any n-qubit circuit $U = U_M \dots U_1$ composed of M elementary gates such that the state $U_t \dots U_1 |0^n\rangle$ is separable for all $1 \leq t \leq M$, it is possible to estimate the probability of measuring 0 on the first qubit classically in polynomial time, up to additive accuracy $1/poly(n)$.*

One method for disproving this conjecture (or rather, showing it to be very unlikely to be true) is to find a class of separable computations for which a 'quantum supremacy' result can be proved (see 3.1). Such a result would state that no classical simulation (to multiplicative or additive) error can exist unless certain complexity theoretic conjectures are false. This is what we had originally attempted for the One Clean Qubit Model without Entanglement.

Chapter 5

Efficient learnability of states

5.1 Classical simulation and efficient learnability

In this chapter we discuss a computational task that, like classical simulation, is classically tractable in some but not all cases. Here we consider whether there is a link between the tractable cases for simulation and this task. The task is that of 'learning' a quantum state. However the notation of learning here is much weaker than required for state tomography where the hypothesis state must be close to the real state (in some metric). Instead we require that using the hypothesis state to create new predictions Probably gives Approximately the Correct answer (PAC). Aaronson proved that remarkably PAC learning only requires a linear number of data points [2], in constrast to full tomography which requires an exponential amount in general.
The catch is that even though one only requires a small amount of data, producing the hypothesis is generally exponentially hard. However, Rocchetto showed that the set of stabiliser states under Pauli measurements can be learned efficiently [78]. Furthermore, the proof of this fact uses tools from the Gottesman–Knill theorem, suggesting that perhaps classical simulation and learnability are related. However, it is not the case that classical simulation implies efficient learnability. It has been shown that if this were true there could be no cryptographic one-way functions [40, 2].
In this chapter we will show that under an addition condition, which we called efficient invertibility, classical simulability does imply learnability. Informally, efficient invertibility requires that given an outcome for a particular measurement it must be possible to (usefully) describe all states that could have given that outcome efficiently. For example, if a stabiliser state gives outcome $+1$ with certainty for a Z_1 measurement, then the stabiliser state must have Z_1 in its stabiliser. This is an efficient and useful description of all the stabiliser states consistent with this measurement outcome.
We show two new examples of efficient learnability and explain how they have this invertibility condition. The first is low Schmidt rank states under measurements that are not too entangling. In the second example we consider states that are efficiently described by a hidden variable

or ontological model, and show that such states are efficiently learnable which we showed is classically simulable in the previous chapter

5.1.1 Definition of PAC learning

In this work we only consider 2 outcome POVMs with outcomes 0 and 1, as this is the typical situation considered in PAC learning. We will therefore only consider classical simulations of quantum circuits with single bit outcomes and no adaptive measurements. Above we described such a circuit as having some input ρ to which a unitary U is applied, and then one qubit (say the first) of the result is measured by the Z_1 operator. Here, in order to make the language more directly comparable to that of PAC learning, we will consider an equivalent formulation. We will think of measuring ρ using the operator $U^\dagger Z_1 U$. This produces the same outcome statistics as the original computation, and in this sense they are equivalent. Therefore if the original scenario is (weakly) classically simulable, then the new situation is too. More formally, the definition of efficiently simulable we will use in this chapter is the following.

Definition 5.1.1. *(Efficient classical simulation) Let $\mathcal{C}_n$ be a class of n-qubit quantum states, and $\mathcal{C} = \bigcup_n \mathcal{C}_n$. Let $\mathcal{M}_n$ be a set of projectors onto the $+1$ outcome for two-outcome measurements on n qubits, and $\mathcal{M} = \bigcup_n \mathcal{M}_n$. We say $\mathcal{C}$ is efficiently classically simulable with respect to $\mathcal{M}$ if:*

- *there exists a canonical classical description of each $E \in \mathcal{M}_n$ of length at most poly(n);*
- *there exists a canonical classical description of each $\rho \in \mathcal{C}_n$ of length at most poly(n);*
- *there exists a classical algorithm L, that takes as input the above descriptions and outputs* $\mathrm{Tr}(E\rho)$ *to additive error ε;*
- *L has runtime polynomial in n and $1/\varepsilon$.*

This notion of simulability is equivalent to weak simulation because the ability to weakly simulate enables one to compute $\mathrm{Tr}(E\rho)$ to $1/\mathrm{poly}(n)$ additive error (by repeating the computation and taking the expectation) and vice versa. However, if instead of $1/\mathrm{poly}(n)$ we required that $\mathrm{Tr}(E\rho)$ is computed to $1/\exp(n)$ additive error, this definition would be that of strong simulation.

Aaronson showed that quantum states are PAC learnable. This describes the following scenario. ρ is some unknown state from $\mathcal{C}$. Let $\{E_i\}_i^m$ be a set of measurements (described by their $+1$ eigenspace projector) drawn according to distribution $\mathcal{D}$ over $\mathcal{M}$. The *data set* contains the probability of getting outcome $+1$ for each measurement, $\{(E_i, \mathrm{Tr}(E_i\rho)\}_i^m$. As long as m, the number of data points, is greater than a particular linear function of n then the following is possible. A hypothesis state σ can be found such that, if a new measurement E is drawn from $\mathcal{M}$, then $|\mathrm{Tr}(E\sigma) - \mathrm{Tr}(E\rho)|$ is small, with good probability. In other words, the hypothesis is Probably Approximately Correct (PAC).

In fact, this theorem does more than just proving such a hypothesis can be found, it gives a very simple description of it. Any state σ which was approximately correct on the training data (in the sense that $|\mathrm{Tr}(E_i\sigma) - \mathrm{Tr}(E_i\rho)|$ was small) will be a good hypothesis. We state this theorem formally below.

Theorem 5.1.2. *(Quantum Occam's Razor [2]) As above, let $\mathcal{C}_n$ be a class of n-qubit quantum states, and $\mathcal{C} = \bigcup_n \mathcal{C}_n$. Let $\mathcal{M}_n$ be a set of two-outcome measurements on n qubits, and $\mathcal{M} = \bigcup_n \mathcal{M}_n$. Let $\{E_1, ..., E_m\}$ be m measurements drawn from the distribution $\mathcal{D}_n$ over $\mathcal{M}_n$. For a given $\rho \in \mathcal{C}_n$ and call $T = \{(E_i, \mathrm{Tr}(E_i\rho))\}_{i\in[m]}$ a training set. Suppose there is a hypothesis state σ such that $|\mathrm{Tr}(E_i\sigma) - \mathrm{Tr}(E_i\rho)| \leq \gamma\varepsilon/7$ for each $i \in [m]$. In words, the hypothesis state is consistent with the data. Then, with probability at least $1-\delta$, the hypothesis state will also be consistent with new data:*

$$\Pr_{E\sim D_n}[|\mathrm{Tr}(E\sigma) - \mathrm{Tr}(E\rho)| \leq \varepsilon] \geq 1 - \delta, \tag{5.1}$$

provided there was enough training data:

$$m \geq \frac{C}{\sigma^2\varepsilon^2}\Big(\frac{n}{\gamma^2\varepsilon^2} log^2 \frac{1}{\gamma\varepsilon} + log\frac{1}{\delta}\Big). \tag{5.2}$$

Even though we know from this theorem we simply need a hypothesis state that fits the training data, it doesn't guarantee that finding such a state will be computationally efficient. The definition of efficiently learnable we use in this work is as follows.

Definition 5.1.3. *(Efficiently learnable) Let $\mathcal{C}$ and $\mathcal{M}$ be as above, and assume there is some canonical efficient classical description of every $E \in \mathcal{M}_n$, so E can be described using $O(poly(n))$ bits* [1]*. We say $\mathcal{C}$ is efficiently learnable with respect to $\mathcal{M}$ if there exists a pair of classical algorithms L_1 and L_2 with the following properties:*

- *(L_1, SDP feasibility algorithm) For every $\rho \in \mathcal{C}_n$, $\eta > 0$, and for every training set $T = \{(E_i, \mathrm{Tr}(E_i\rho))\}_{i\in[m]}$, L_1 outputs a poly(n) bit classical description of a hypothesis state σ that satisfies the following approximate feasibility problem*

$$\begin{aligned} |\mathrm{Tr}(E_i\sigma) - \mathrm{Tr}(E_i\rho)| \leq \eta \quad &\textit{for all } \ i \in [m], \\ \sigma \succeq 0, \\ \mathrm{Tr}(\sigma) = 1. \end{aligned} \tag{5.3}$$

  L_1 runs in time $poly(n, m, 1/\eta)$.

- *(L_2, simulation algorithm) For every n qubit σ which is an output of L_1 and for every $E \in \mathcal{M}_n$, L_2 computes $\mathrm{Tr}(E\sigma)$. L_2 runs in time $poly(n)$.*

[1] It is natural to assume one exists, because otherwise it would take exponential time for the experimenter to even describe the measurement they will perform.

We define $\mathcal{S}$ to be the set of all possible hypothesis states σ that satisfy Eq. 5.1.3. In the language of learning theory $\mathcal{S}$ is the hypothesis state space of the quantum state learning problem. Note that the definition above ensures that all hypothesis states have efficient classical descriptions.
We remark that our notion of efficient learnability is slightly different from previous definitions [2, 78]. In our definition we emphasise that efficient learnability does not only require an efficient way to produce a description of the hypothesis state σ, but also that this description can be used to efficiently compute $\mathrm{Tr}(\sigma E)$ for all $E \in \mathcal{M}$. In the language of classical simulation that was introduced at the beginning of the subsection, this amounts to efficiently simulating $\mathcal{S}$ under the measurement set $\mathcal{M}$. The motivation for this extra requirement is that efficient learnability ought to imply that a classical computer can efficiently produce a prediction for a new measurement of the state.

5.2 A condition under which classical simulability implies efficient learnability

In this subsection we discuss why classical simulation on its own is not enough to imply learnability, but provide an extra condition, which we called efficient invertibility, which does allow this implication. First we explain why classical simulabilty gives some, but not all, of the conditions needed for learnability.
If an efficient classical simulation algorithm exists for $\mathcal{C}$ with respect to $\mathcal{M}$, that guarantees that an efficient description exists for all $\rho \in \mathcal{C}$ and that, given the efficient description, it is possible to compute $\mathrm{Tr}(\rho E)$ for all $E \in \mathcal{M}$. Recall that these properties are required in the definition of efficient learnability (in the case where $\mathcal{S} = \mathcal{C}$). This is called L_2 or the simulation algorithm in Def. 5.1.3. However, this does not guarantee that the SDP feasibility algorithm (also required in the definition) exists. In fact, Aaronson provides a counterexample [2]. If all classically simulable circuits are (even quantumly) learnable, that implies there are no cryptographic one-way functions (safe from a quantum attack).
In this section we explain that an extra property is required on top of simulability for this implication. To illustrate this, we will examine the efficient learning algorithm in the example of stabiliser states under Pauli measurements.
The algorithm follows this procedure (elaborated in [78]):

1. For each training set data point $\{E_i, \mathrm{Tr}(E_i\rho)\}$, let $\mathcal{S}_i$ be the set of stabiliser states consistent with this measurement outcome: $\mathcal{S}_i = \{\phi \in \mathcal{C} : \mathrm{Tr}(E\phi) = \mathrm{Tr}(E\rho)\}$. It is possible to characterise states in this set in terms of their stabilisers.

2. It is also possible to characterise states in the intersection $\bigcap \mathcal{S}_i$ in terms of their stabilisers. States in this set are consistent with all data points. Choose one such state σ.

3. Use the Gottesman-Knill Theorem to make predictions using σ for future Pauli measurements.

While the classical simulability of stabiliser states under Pauli measurements was necessary for the final step, we see that there is another property that is important: the ability to efficiently characterise all states consistent with the data. In other words, to describe $\mathcal{S}_i$ efficiently, and from this, efficiently compute the intersection of the sets. Then finally, it must be possible to efficiently describe a particular state σ in this set. In particular, that efficient description must be the one that the classical simulation algorithm uses. In this case, the final description of σ is in terms of its stabilisers, which allows us to use the Gottesman-Knill simulation.
We call this condition the *efficient invertibility condition*, as we require to invert $\mathrm{Tr}(E_i\rho)$ (i.e. to find the set of ϕ that matches the expectation for each training measurement E_i). Any state in the intersection corresponds to a solution to the SDP feasibility problem.
Classical simulation is related to the conditions listed before. If $\mathcal{C}$ is classically simulable with respect to $\mathcal{M}$, that guarantees that an efficient description exists for all $\rho \in \mathcal{C}$, and that given the efficient description, it is possible to compute $\mathrm{Tr}(\rho E)$. A trivial learning algorithm based on classical simulation would then take $\mathcal{S} = \mathcal{C}$, compute the expectation value of all the elements in the hypothesis set and return any of the states that is consistent with all the measurements. But in general $|\mathcal{C}| = |\mathcal{S}| = O(\exp(n))$ and thus this algorithm would be terribly inefficient. In the stabiliser example it was important that, aside from $\mathrm{Tr}(\sigma E)$ being efficient to calculate in those cases, it is was also efficient to characterise all states ϕ that had a particular value of $\mathrm{Tr}(\phi E)$. In the next subsection we provide another example where the invertibility condition is crucial.
We will now generalise the algorithm developed for stabiliser states, making allowance for the hypothesis state to only be consistent with the data to additive error η (as per definition 5.1.3). This learning algorithm can be schematised as follows:

1. characterise all states in $\mathcal{S}_i^{\eta'} = \{\phi \in \mathcal{C} : |\mathrm{Tr}(E_i\rho) - \mathrm{Tr}(E_i\phi)| \leq \eta'\}$ (with $0 < \eta' \leq \eta$) in time polynomial in n and $1/\eta$ and,

2. produces an efficient description of a state σ in the intersection of all $\mathcal{S}_i^{\eta'}$, $i \in [0,m]$, in time polynomial in m, n and $1/\eta$,

3. and then uses this description to compute $\mathrm{Tr}(\sigma E)$, for any $E \in \mathcal{M}$, in time polynomial in n.

If all the conditions hold then $\mathcal{C}$ is efficiently learnable with respect to $\mathcal{M}$. This may not be the only condition under which efficient learning is possible. However, it appears to be natural as it holds for the three known examples of efficient learning, namely stabilisers and the examples we provide in the following subsections.

5.2.1 Slightly entangled states are efficiently learnable

Quantum computations in which the state of the computer does not become entangled are classically simulatable [51, 88]. In fact, even if the states are allowed a 'slight' amount of entanglement, as defined below, it is still possible to efficiently simulate their evolution. Here we will show that such states are also efficiently learnable under suitable measurements.

Definition 5.2.1. *(Schmidt decomposition) The Schmidt decomposition of a state $|\psi\rangle$ with respect to the $A|B$ partition is*

$$|\psi\rangle_{AB} = \sum_{i=1}^{r} \sqrt{\lambda_i}|a_i\rangle_A|b_i\rangle_B, \tag{5.4}$$

where the $|a_i\rangle$ are orthonormal, as are the $|b_i\rangle$. Assume $\lambda_i \neq 0$ for all i. The Schmidt rank of the $A|B$ partition is r.

See [67] for the proof that a Schmidt decomposition exists for all states and all partitions, and that the Schmidt rank is unique.
Note that if the Schmidt rank is 1, the A and B subsystems are not entangled. Generally though the Schmidt rank can be exponentially large in the dimension of the subsystems. This motivates the following definition.

Definition 5.2.2. *(Slightly entangled states) A class of states is called slightly entangled if, for every n qubit state $|\psi\rangle$ in the class and every partition $A|B$, the Schmidt rank is bounded by a polynomial. This class is called L-entangled if the polynomial bounding the Schmidt rank is $L(n)$.*

Ref [88] shows that if the state of a quantum computation is L-entangled through out the computation, this computation can be classically simulated. Ref [93] gives an example of circuits that have this property. They show that a computation which is $log(n)$ depth, has bounded range interactions, and has product inputs has a state during the computation that is L-entangled and hence classically simulable. Ref [49] generalised this to show the following. Suppose $\mathcal{C}$ is a circuit comprised on 1 and 2 qubit gates. Let D_i be the number of gates that either act of line i, or act across that line (i.e D_i is the number of two qubit gates that act on lines j and k with $j \leq i \leq k$). Let $D = max_i D_i$. Then if D is bounded by a polynomial and the input to the circuit $\mathcal{C}$ is a product state, then the quantum state during the computation is slightly entangled, and hence $\mathcal{C}$ is efficiently classically simulable.
We will now define low Schmidt rank measurements as the measurements induced by circuits in the above form.

Definition 5.2.3. *(Low Schmidt rank measurements) Suppose U is a n qubit unitary comprised of 1 and 2 qubit gates, with $D = max_i D_i$ as above, and suppose D is bounded by a polynomial*

in n. Then we call the measurement $U^\dagger Z_i U$ a low Schmidt rank measurement, where Z_i is a Z measurement of the ith line.

This definition is motivated by noting that if such a measurement were performed on a slightly entangled state, the result would be classically simulable in the sense of 5.1.1. The name refers to the property that such a measurement keeps the Schmidt rank of a slightly entangled state low. It does not imply that these measurements are low rank.

Theorem 5.2.4. *The class of L-entangled states is efficiently learnable under low Schmidt rank measurements.*

Proof. We will first give an efficient parametric description of an unknown L-entangled state. This description involves $2nL^2$ complex parameters at most and any state in that form is necessarily an L-entangled state.
The method of deriving this description is described in detail on page 5 of Ref [49] where a proof of it's correctness is also given. We will omit that proof here, but explain the process.
Suppose $|\psi\rangle$ is a L-entangled n qubit state, and consider the partition of the qubits $1|2...n$. Use the Schmidt decomposition to write

$$|\psi\rangle = \sum_j |a_j\rangle|\nu_j\rangle, \tag{5.5}$$

where the Schmidt coefficients are absorbed into $|a_j\rangle$, so they are orthogonal and subnormalised while $|\nu_j\rangle$ are orthonormal. Let $|a_j\rangle = \alpha_j^0|0\rangle + \alpha_j^1|1\rangle$. For $|\psi\rangle$ to be properly normalised, $\sum_j |\alpha_j^0|^2 + |\alpha_j^1|^2 = 1$.
Now consider the partition $12|3...n$ of $|\psi\rangle$. Let $|\nu_k\rangle$ be the orthonormal Schmidt vectors on qubits $3,...,n$. Then

$$|\nu_j\rangle = \sum_k |b_{jk}\rangle|\eta_k\rangle. \tag{5.6}$$

The $|b_{jk}\rangle$s are subnormalised. Let $|b_{jk}\rangle = \beta_{jk}^0|0\rangle + \beta_{jk}^1|1\rangle$. Then because $|\nu\rangle$ is normalised and all $|\eta_k\rangle$ are orthonormal, $\sum_j |\beta_{jk}^0|^2 + |\beta_{jk}^1|^2 = 1$, for all j.
Continuing in this way, we can write all L-entangled states parametrically as

$$|\psi\rangle = \sum_{j,k,l..p} |a_j\rangle|b_{jk}\rangle|c_{kl}\rangle...|k_p\rangle, \tag{5.7}$$

with normalisation equations as above. Each index sums at most to L, and hence there are at most $2nL^2$ complex parameters in this description.
Now we will use this description to show such states are learnable under low Schmidt rank measurements. Suppose the measurements in the training set are $E_1,...,E_m$, and the training data is $\{\mathrm{Tr}(|\psi\rangle\langle\psi|E_i)\}_i$ The steps of learning algorithm follow the same pattern as the previous examples.

1. Let E_i be the projector onto the positive subspace of the measurement $U^\dagger Z_j U$. Then it is possible to find the L-entangled states consistent with the data point $(E_i, \mathrm{Tr}(|\psi\rangle\langle\psi|E_i)$ by first evolving the state in Eq. 5.7 by U. This is possible to do efficiently by the method given in Ref [**?**]. Then it is also possible to find efficiently find the expression for the probability of outcome $+1$ when line j of this state is measured in the Z basis. This expression is a polynomial in the parameters. Equating that with $\mathrm{Tr}(|\psi\rangle\langle\psi|E_i)$ gives a polynomial equation that $|\psi\rangle$ needs to satisfy in order to be consistent with the data point.

2. To find a $|\sigma\rangle$ consistent with all the data, all the above polynomial equations must be solved simultaneously. We do not have to find the correct solution (the one corresponding to $|\psi\rangle$). Any solution wlll yield a description of a hypothesis state in the form of Eq. 5.7.

3. Given this description, it is efficient to compute the measurement outcomes of $|\sigma\rangle$ for any low Schmidt rank measurement.

□

5.2.2 Efficient ontological models are efficiently learnable

In this section we will show that if a set of states can be described efficiently in the ontological models framework, then they are both efficiently simulable and efficiently learnable. The ontological models framework was introduced in Ref [46] and is a general framework for theories that reproduce the statistics of quantum mechanics.

Definition 5.2.5. *(Ontological model) Let $\mathcal{C}$ be a set of quantum states, and $\mathcal{M}$ be a set of measurements previously*[2]*. Then an ontological model with underlying ontic state space Λ_n describes the states in $\mathcal{C}_n$ with respect to $\mathcal{M}_n$, if the following conditions hold. There exists a function $f : \Lambda_n \times \mathcal{M}_n \to [0,1]$; for all $\rho \in \mathcal{C}_n$, and there exists a probability distribution p_ρ over Λ_n such that, for all $\rho \in \mathcal{C}_n$ and $E \in \mathcal{M}_n$, $\mathrm{Tr}(\rho E) = \sum_{\lambda\in\Lambda_n} p_\rho f(\lambda, E)$.*

Essentially, the ontic state space represents the true (potentially hidden) state of a system in the quantum state ρ. In this model, the system is thought to be in some ontic state $\lambda \in \Lambda$ with probability $p_\rho(\lambda_n)$. There is a function that assigns the probability of outcome $+1$ for measurement E as $f(\lambda, E)$. For a deterministic hidden variable theory, $f(\lambda, E)$ is always 0 or 1. To recover the usual quantum formalism instead, let Λ_n be the n qubit quantum state space and p_ρ be a delta Dirac function probability centred on ρ. Finally, let f assign Born Rule probabilities. Hence, this model is rich enough to describe the usual quantum formalism as well as hidden variable descriptions of quantum mechanics.
We will now add a very natural extra restriction that will make the ontological models we consider classically simulable.

[2]We will assume that a full context is prespecified for each of the measurements in $\mathcal{M}$. This allows us to talk about ontological models that are contextual without ambiguity.

Definition 5.2.6. *(Efficient ontological model (EOM)) An efficient ontological model for $\mathcal{C}$ under $\mathcal{M}$ is an ontological model for which*

- $|\Lambda_n| = O(poly(n))$,
- *there exists a classical algorithm for evaluating $f(\lambda, E)$ and $p_\rho(\lambda)$, for all $\rho \in \mathcal{C}_n$, $\lambda \in \Lambda_n$, and $E \in \mathcal{M}_n$, in runtime polynomial in n.*

Theorem 5.2.7. *If there exists an EOM for the set of states $\mathcal{C}$ and measurements $\mathcal{M}$, then it is possible to classically simulate $\mathcal{C}$ with respect to $\mathcal{M}$ in the sense of Definition 5.1.1.*

Proof. Let ρ a state in $\mathcal{C}$ and E be a measurement in $\mathcal{M}$. The probability distribution $p_\rho(\lambda)$ can be computed for each $\lambda \in \Lambda$. $|\Lambda| = O(\text{poly}(n))$ and so this can be done for every λ efficiently, and hence $p_\rho(\lambda)$ can be sampled from efficiently. To efficiently compute $\text{Tr}(\rho E)$, (1) sample λ' from p_ρ, (2) compute $f(\lambda', E)$, (3) repeat this algorithm polynomially many times to estimate $\text{Tr}(\rho E)$ to sufficient accuracy. □

So far we have only considered probability distributions p_ρ over Λ that represent the state of a quantum system ρ, and therefore produce the outcome statistics of ρ. However, it is possible to consider more general probability distributions over Λ, which are what we'll call *preparations*. The outcome statistics for these will generally not correspond to any state or obey the Born rule. However, they will be useful to consider for us. This is because if the states and measurements being learned are described by an EOM, instead of producing a hypothesis quantum state σ and using that to predict future outcome statistics, it will be simpler to produce a hypothesis preparation instead. The outcomes predicted from such a preparation will in fact be a good prediction with high probability, in the usual learning sense. We will now formalise this discussion.

Definition 5.2.8. *(Preparation) For an ontological model with ontic state space Λ, a preparation is a probabilistic mixture over Λ. Suppose that the probability distribution is given by $p : \Lambda \to [0,1]$. Identify the preparation with p.*

Theorem 5.2.9. *EOMs are efficiently learnable.*

Proof. Suppose that there is an EOM with ontic state space Λ ($|\Lambda| = O(n)$) with measurement outcome assigning function $f : \Lambda \times \mathcal{M} \to \{0,1\}$. Let p be the unknown preparation to learn and suppose the training data is $T = \{(E_i, d_i = \sum_\lambda p(\lambda) f(\lambda, E_i))\}_{i \in [m]}$. Then let the hypothesis state space be all possible preparations on Λ.

1. The preparations (exactly) compatible with the data point $(E_i, d_i = \sum_\lambda p(\lambda) f(\lambda, E_i))$ are $\{q : d_i = \sum_\lambda q(\lambda) f(\lambda, E_i))\}$. It is efficient to write this condition in full, as the number of terms in the equation is polynomial in n and computing $f(\lambda, E_i)$ is also efficient, by assumption. Also add the constraint that $\sum_\lambda q(\lambda) = 1$.

2. Each of these m constraint equations are linear, and involve $|\Lambda| = O(n)$ unknowns. It is possible to find a solution in time polynomial in m and n. There may not be a unique solution, but any solution (for which q is a probability distribution) is sufficient.

3. Given the hypothesis preparation q, it is efficient to compute $\sum_\lambda q(\lambda) f(\lambda, E)$ for any $E \in \mathcal{M}$.

□

This theorem shows that preparations are efficiently learnable in the sense of Definition 5.1.3. However, the PAC learning theorem (Theorem 5.1.2) does not apply in this instance, as not all preparations are representations of quantum states. We require a generalisation of the PAC theorem in this case in order to justify the above procedure as 'learning'. We provide such a generalisation:

Theorem 5.2.10. *(Occam's razor for EOMs) EOMs are PAC learnable, in the sense of Theorem 5.1.2.*

We will prove this theorem in the rest of the chapter.

A set of functions is PAC learnable as long the functions are not too 'flexible'. This is what allows one to predict future values of the function from a small amount of training data. The flexibility of a set of functions is quantified by the fat-shattering dimension.

Definition 5.2.11. *(Fat-shattering dimension) Let $\mathcal{M}$ be a sample space, and $\mathcal{V}$ be a set of functions that map elements of $\mathcal{M}$ to numbers in $[0,1]$. We say a set $\{E_1, ..., E_k\} \subset \mathcal{M}$ is γ-fat shattered by $\mathcal{V}$ if there exists real numbers $\alpha_1, ..., \alpha_k$ such that all $B \subseteq \{1, ...k\}$, there exists $g \in \mathcal{V}$ such that for all $i \in \{1, ..., k\}$,*

- *if $i \notin B$ then $g(E_i) \leq \alpha_i - \gamma$, and*
- *if $i \in B$ then $g(E_i) \geq \alpha_i + \gamma$.*

The γ-fat-shattering dimension of $\mathcal{V}$, $fat_\mathcal{V}(\gamma)$, is the maximum k such that some $\{E_1, ..., E_k\}$ is γ-fat-shattered by $\mathcal{V}$.

In the quantum case, let $\mathcal{C}$ be a class of quantum states, and $\mathcal{M}$ be a set of measurements. The function $\mathrm{Tr}(\cdot\rho) : \mathcal{M} \to [0,1]$ takes measurements in $\mathcal{M}$ and outputs the probability of getting outcome $+1$ when measuring ρ. $\mathcal{V}$ is the set of these functions, for all $\rho \in \mathcal{C}$.
For the preparations of an EOM case, if $\mathcal{P}$ is the set of preparations, $\mathcal{V}$ is the set of functions in the form $\sum_\lambda p(\lambda) f(\lambda, \cdot)) : \mathcal{M} \to [0,1]$, for $p \in \mathcal{P}$.
We now state a result that links the learnability of a class of functions with its fat-shattering dimension.

Theorem 5.2.12. *(From Ref [7]) Define $\mathcal{M}$ and $\mathcal{V}$ as above, and let $\mathcal{D}$ be a probability distribution over $\mathcal{M}$. Fix an element $g \in \mathcal{V}$, as well as error parameters $\varepsilon, \eta, \gamma > 0$, with $\gamma > \eta$. Suppose $\{E_1, ..., E_m\}$ are drawn from $\mathcal{P}$ independently according to $\mathcal{D}$. Let $T = \{(E_i, g(E_i))\}_i^m$ be the training set. Let $h \in \mathcal{V}$ be a hypothesis consistent with the training set: $|h(E_i) - g(E_i)| \leq \eta$ for $i \in [1, m]$. Then there exists positive constant K such that if*

$$m \geq \frac{K}{\varepsilon}\Big(fat_{\mathcal{V}}\big(\frac{\gamma-\eta}{8}\big) log^2\big(\frac{fat_{\mathcal{V}}(\gamma-\eta/8)}{(\gamma-\eta)\varepsilon}\big) + log\frac{1}{\delta}\Big), \tag{5.8}$$

then with probability at least $1-\delta$,

$$Pr_{E\sim\mathcal{D}}[|h(E) - g(E)| > \gamma] \leq \varepsilon. \tag{5.9}$$

Bounding the fat-shattering dimension for preparations of EOMs will allow us to bound the expression in Equation 5.8 in this case. First we will prove a result about random access codes using preparations of an EOM. The equivalent theorem for the quantum and classical state was proved in [6].

Theorem 5.2.13. *Let k and n be positive integers with $k > n$. For a k-bit string $y = y_1...y_k$, let p_y be a preparation for an EOM. Suppose there exists measurements in $\mathcal{M}$ such that, for all $y \in \{0,1\}^k$ and $i \in \{1, ..., k\}$,*

- *if $y_i = 0$ then the probability of outcome $+1$ for measurement E_i is greater than p,*
- *if $y_i = 1$ then the probability of outcome -1 for measurement E_i is greater than p.*

Then $O(log(n)) \geq (1 - H(p))k$, where H is the binary entropy function.

This theorem bounds the size of strings that can be encoded in a preparation. The measurement E_i returns the correct value of y_i with probability p.

Proof. Let $Y = Y_1...Y_k$ be the random chosen uniformly at random from $\{0,1\}^k$. For a preparation p_y with $y \sim Y$, let X be a random variable that chooses an ontic state $\lambda \in \Lambda$ with probability p_y. Let Z_i be the random variable that records the result of measuring a preparation $p_{y\sim Y}$ with E_i, and let $Z = Z_i...Z_m$.
The mutual information of Y and X is bounded by the number of bits needed to specify $\lambda \sim X$.

$$I(Y:X) \leq S(X) \leq \log(|\Lambda|) = O(\log(n)) \tag{5.10}$$

We will now bound the mutual information in the other direction.

$$I(Y:X) \leq S(Y) - S(Y|X) = k - S(Y|X), \tag{5.11}$$

and,

$$S(Y|X) \leq S(Y|Z) \leq \sum_{i=1}^{k} S(Y_i|Z) \leq \sum_{i=1}^{k} S(Y_i|Z_i). \tag{5.12}$$

Because knowing $z_i \sim Z_i$ gives us the correct value of $y_i \sim Y$ with probability greater than or equal to p, $S(Y_i|Z_i) \leq H(p)$. Hence, $O(\log(n)) \leq k(1-H(p))$.

□

This theorem is necessary to prove the fat-shattering theorem, which is analogous to Theorem 2.6 in [2].

Lemma 5.2.14. *Let k, n and $\{p_y\}$ be as in Theorem 5.2.13. Suppose there exists $E_1, ..., E_k \in \mathcal{M}$ and real numbers $\alpha_i, ..., \alpha_k$, such that for all $y \in 0,1^k$ and $i \in \{1, ..., k\}$,*

- *if $y_i = 0$ then the probability of outcome $+1$ for measurement E_i is greater than $\alpha_i - \gamma$,*
- *if $y_i = 1$ then the probability of outcome $+1$ for measurement E_i is less than $\alpha_i + \gamma$.*

Then $O(log(n))/\gamma^2 = O(k)$

The proof is essentially the same as for Theorem 2.6 in [2]. Interpreting k as the fat-shattering dimension gives us our result.

Corollary 5.2.15. *For all $\gamma > 0$, the fat-shattering dimension of preparations of an EOM is $O(n/\gamma^2)$.*

Combining this with Theorem 5.2.12 shows that preparations of an EOM are PAC learnable.

5.3 Remarks

Efficiently PAC learnable and classically simulable are two related notions of classicality. While neither implies the other, it is fertile to understand what extra conditions are necessary for an implication. Doing so allowed us to identify two new classes of states that are efficiently PAC learnable.

Chapter 6

Conclusion

It is still very much an open question what the source of the quantum advantage is. In this thesis we identified a few different quantum resources that where each necessary for a quantum speedup in particular situations. For two of these examples we provided some suggestive evidence that resources are not just necessary but are in fact being utilised by the quantum computer.

We proved the extended Gottesman–Knill theorem. This showed that T gates are in fact crucial to quantum computers. The power of a quantum computer is proportional to the number of T gates and not the other resources such as the number of qubits. We showed this by showing that any quantum algorithm with t number of T gates can be simulated with just those T gates and t qubits, as well as Clifford gates and classical polynomial time computation. This suggests that T gates are an important resource being used by quantum computers.

We also showed that in the One Clean Qubit Model that entanglement necessary. This result is surprising because it is not known whether entanglement is necessary for mixed state quantum computers to have an advantage, and for this computer it was suspected that it was not. That is because this model has no entanglement between the only qubit that begins pure and the noisy register. As the clean qubit is the one that is ultimately measured, we might have suspected that entanglement across this cut would be necessary to create correlations that cannot be classically simulated in this model, which is not the case. The unlikeliness of this simulation suggests that entanglement truly is an important factor in the quantum advantage of the One Clean Qubit Model, and quantum computing more generally.

In the future we hope to see progress made on other methods to attack the question "what resources are crucial for quantum computers?". For example, instead of only seeing that a resource is necessary for a speedup, we would like to see a mechanism that explains how this property is useful. A big step in this direction came from Brayi, Gosset and Koenig [17]. They proved that quantum computers could solve a particular mathematical problem in constant depth

while a classical computer cannot. The method used Bell's theorem to show that a computer based on a local theory like classical mechanics cannot fully replicate the correlations of this quantum computer. This suggests that it is the speed with which entanglement allows quantum correlations to spread that matters. While the result is suggestive, it is not conclusive because it is only something that holds for a particular problem and the speed up it leads to is measly in this case. It would make a very strong case for entanglement's importance in quantum computing if it is shown that large speedups can also be achieved by fast quantum correlations.

On the other hand, there many be results in the future that strong suggest entanglement is not crucial. For example, perhaps mixed state quantum computers don't all need entanglement for an advantage after all? Or if one could show that even with very small amounts of entanglement in a pure computer that one can achieve quantum speedups, that would show entanglement is not crucial because it cannot be being used as a resource in those amounts. Van den Nest proved a tantalisingly similiar result. He showed that a circuit family can have *decreasing* amounts of entanglement and still have an advantage [85]. However, in that case the amount of entanglement was decreasing but not truly small. To convincingly show entanglement isn't important one would need to show that even exponentially vanishing amounts of entanglement are enough for a speedup.

Understanding entanglement's true role in quantum computation is an important direction for future research in quantum computation. The goal, though, should be to apply what we learn from computation back to physics. What do quantum computers have to tell us about clockwork of the universe?

Bibliography

[1] Aaronson, S. (2005). Ten semi-grand challenges for quantum computing theory.

[2] Aaronson, S. (2007). The learnability of quantum states. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 463, pages 3089–3114. The Royal Society.

[3] Aaronson, S. and Arkhipov, A. (2011). The computational complexity of linear optics. In *Proceedings of the 43rd annual ACM symposium on Theory of computing - STOC '11*, page 333, New York, New York, USA. ACM Press.

[4] Aaronson, S. and Gottesman, D. (2004a). Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328.

[5] Aaronson, S. and Gottesman, D. (2004b). Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328.

[6] Ambainis, A., Nayak, A., Ta-Shma, A., and Vazirani, U. (2002). Dense quantum coding and quantum finite automata. *Journal of the ACM (JACM)*, 49(4):496–511.

[7] Anthony, M. and Bartlett, P. (1995). Function learning from interpolation. In *European Conference on Computational Learning Theory*, pages 211–221. Springer.

[8] Arora, S. and Barak, B. (2009). *Computational complexity : a modern approach*. Cambridge University Press.

[9] Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J. C., Barends, R., Biswas, R., Boixo, S., Brandao, F. G., Buell, D. A., Burkett, B., Chen, Y., Chen, Z., Chiaro, B., Collins, R., Courtney, W., Dunsworth, A., Farhi, E., Foxen, B., Fowler, A., Gidney, C., Giustina, M., Graff, R., Guerin, K., Habegger, S., Harrigan, M. P., Hartmann, M. J., Ho, A., Hoffmann, M., Huang, T., Humble, T. S., Isakov, S. V., Jeffrey, E., Jiang, Z., Kafri, D., Kechedzhi, K., Kelly, J., Klimov, P. V., Knysh, S., Korotkov, A., Kostritsa, F., Landhuis, D., Lindmark, M., Lucero, E., Lyakh, D., Mandrà, S., McClean, J. R., McEwen, M., Megrant, A., Mi, X., Michielsen, K., Mohseni, M., Mutus, J., Naaman, O., Neeley, M., Neill, C., Niu, M. Y., Ostby, E., Petukhov, A., Platt, J. C., Quintana, C., Rieffel, E. G., Roushan, P., Rubin, N. C., Sank, D., Satzinger, K. J., Smelyanskiy, V., Sung, K. J., Trevithick, M. D., Vainsencher, A., Villalonga, B., White, T., Yao, Z. J., Yeh, P., Zalcman, A., Neven, H., and Martinis, J. M. (2019). Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510.

[10] Bell, J. S. (1964). On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195.

[11] Bermejo-Vega, J., Delfosse, N., Browne, D. E., Okay, C., and Raussendorf, R. (2017). Contextuality as a resource for models of quantum computation on qubits. *Physical Review Letters*, 119(120505).

[12] Bohm, D. (1952). A suggested interpretation of the quantum theory in terms of "hidden" variables. I. *Physical Review*, 85(2):166–179.

[13] Boixo, S., Isakov, S. V., Smelyanskiy, V. N., Babbush, R., Ding, N., Jiang, Z., Bremner, M. J., Martinis, J. M., and Neven, H. (2016). Characterizing Quantum Supremacy in Near-Term Devices. *arXiv:1608.00263*.

[14] Bouland, A., Fefferman, B., Nirkhe, C., and Vazirani, U. (2018). Quantum Supremacy and the Complexity of Random Circuit Sampling. *arXiv: 1803.04402*.

[15] Bouland, A., Fitzsimons, J. F., and Koh, D. E. (2017). Quantum Advantage from Conjugated Clifford Circuits.

[16] Bravyi, S. (2005). Lagrangian representation for fermionic linear optics. *Quantum Information & Computation*, 5(3):216–238.

[17] Bravyi, S., Browne, D., Calpin, P., Campbell, E., Gosset, D., and Howard, M. (2018a). Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3(181).

[18] Bravyi, S. and Gosset, D. (2016). Improved Classical Simulation of Quantum Circuits Dominated by Clifford Gates. *Physical Review Letters*, 116(25):250501.

[19] Bravyi, S., Gosset, D., and Koenig, R. (2018b). Quantum advantage with shallow circuits. *Science*, 362(6412):308–311.

[20] Bravyi, S., Gosset, D., Koenig, R., and Tomamichel, M. (2019). Quantum advantage with noisy shallow circuits in 3D. In *IEEE 60th Annual Symposium on Foundations of Computer Science*, pages 995–999.

[21] Bravyi, S. and Kitaev, A. (2004). Universal Quantum Computation with ideal Clifford gates and noisy ancillas.

[22] Bravyi, S., Smith, G., and Smolin, J. A. (2016). Trading Classical and Quantum Computational Resources. *Physical Review X*, 6(2):021043.

[23] Bremner, M. J., Jozsa, R., and Shepherd, D. J. (2010). Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2126):459–472.

[24] Bremner, M. J., Jozsa, R., and Shepherd, D. J. (2011). Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2126):459–472.

[25] Bremner, M. J., Montanaro, A., and Shepherd, D. J. (2016a). Achieving quantum supremacy with sparse and noisy commuting quantum computations. *arXiv:1610.01808*.

[26] Bremner, M. J., Montanaro, A., and Shepherd, D. J. (2016b). Average-Case Complexity Versus Approximate Simulation of Commuting Quantum Computations. *Physical Review Letters*, 117(8):080501.

[27] Bremner, M. J., Montanaro, A., and Shepherd, D. J. (2016c). Average-case complexity versus approximate simulation of commuting quantum computations. *Physical review letters*, 117(8):080501.

[28] Brod, D. J. (2016). Efficient classical simulation of matchgate circuits with generalized inputs and measurements. *Physical Review A*, 93(6).

[29] Bu, K. and Koh, D. E. (2019). Efficient classical simulation of Clifford circuits with nonstabilizer input states. *Physical Review Letters*, 123(170502).

[30] Cade, C. and Montanaro, A. (2018). The quantum complexity of computing schatten p-norms. In *13th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. `arXiv:1706.09279`.

[31] Dankert, C., Cleve, R., Emerson, J., and Livine, E. (2009). Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80(1):012304.

[32] Danos, V., Kashefi, E., and Panangaden, P. (2007). The measurement calculus. *Journal of the ACM*, 54(2):8–es.

[33] Datta, A., Flammia, S. T., and Caves, C. M. (2005). Entanglement and the power of one qubit. *Physical Review A*, 72(4):042316.

[34] Datta, A. and Vidal, G. (2007). Role of entanglement and correlations in mixed-state quantum computation. *Physical Review A*, 75(4):042310.

[35] Deutsch, D. and Jozsa, R. (1992). Rapid Solution of Problems by Quantum Computation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 439(1907):553–558.

[36] Divincenzo, D. P. and Terhal, B. M. (2005). Fermionic Linear Optics Revisited. *Foundations of Physics*, 35(12).

[Fefferman and Umans] Fefferman, B. and Umans, C. On the Power of Quantum Fourier Sampling.

[38] Frembs, M., Roberts, S., and Bartlett, S. (2018). Contextuality as a resource for measurement-based quantum computation beyond qubits. *New Journal of Physics*, 20.

[39] Fujii, K., Kobayashi, H., Morimae, T., Nishimura, H., Tamate, S., and Tani, S. (2015). Power of quantum computation with few clean qubits. *arXiv preprint arXiv:1509.07276*.

[40] Goldreich, O., Goldwasser, S., and Micali, S. (1986). How to construct random functions. *Journal of the ACM (JACM)*, 33(4):792–807.

[41] Gottesman, D. (1997). Stabilizer codes and quantum error correction. *arXiv preprint quant-ph/9705052*.

[42] Gottesman, D. (1998). Theory of fault-tolerant quantum computation. *Physical Review A*, 57(1):127–137.

[43] Gross, D., Flammia, S. T., and Eisert, J. (2009). Most Quantum States Are Too Entangled To Be Useful As Computational Resources. *Physical Review Letters*, 102(190501).

[44] Haferkamp, J., Hangleiter, D., Bouland, A., Fefferman, B., Eisert, J., and Bermejo-Vega, J. (2019). Closing gaps of a quantum advantage with short-time Hamiltonian dynamics.

[45] Hangleiter, D., Bermejo-Vega, J., Schwarz, M., and Eisert, J. (2017). Anti-concentration theorems for schemes showing a quantum speedup. *arXiv: 1706.03786*.

[46] Harrigan, N. and Spekkens, R. W. (2010). Einstein, incompleteness, and the epistemic view of quantum states. *Foundations of Physics*, 40(2):125–157.

[47] Hebenstreit, M., Jozsa, R., Kraus, B., Strelchuk, S., and Yoganathan, M. (2019). All Pure Fermionic Non-Gaussian States Are Magic States for Matchgate Computations. *Physical Review Letters*, 123(8).

[48] Howard, M., Wallman, J., Veitch, V., and Emerson, J. (2014). Contextuality supplies the 'magic' for quantum computation. *Nature*, 510(7505):351–355.

[49] Jozsa, R. (2006). On the simulation of quantum circuits. *arXiv preprint quant-ph/0603163*.

[50] Jozsa, R. and den Nest, M. V. (2014). Classical simulation complexity of extended Clifford circuits. *Quantum Information & Computation*, 14(7&8).

[51] Jozsa, R. and Linden, N. (2003). On the role of entanglement in quantum-computational speed-up. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 459(2036):2011–2032.

[52] Jozsa, R. and Miyake, A. (2008). Matchgates and classical simulation of quantum circuits. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 464(2100):3089–3106.

[53] Jozsa, R. and Van den Nest, M. (2013). Classical simulation complexity of extended Clifford circuits.

[54] Knill, E. (2001). Fermionic Linear Optics and Matchgates.

[55] Knill, E. and Laflamme, R. (1998). Power of one bit of quantum information. *Physical Review Letters*, 81(25):5672.

[56] Koh, D. E. (2015). Further extensions of Clifford circuits and their classical simulation complexities. *Quantum Information & Computation*, 17(3&4):0262–0282.

[57] Mann, R. L. and Bremner, M. J. (2017). On the Complexity of Random Quantum Computations and the Jones Polynomial. *arXiv: 1711.00686*.

[58] Markov, I. L. and Shi, Y. (2008). Simulating quantum computation by contracting tensor networks. *SIAM Journal on Computing*, 38(3):963–981.

[59] Mermin, N. D. (1990). Simple unified form for the major no-hidden-variables theorems. *Physical Review Letters*, 65(27):3373–3376.

[60] Mermin, N. D. (1993). Hidden variables and the two theorems of John Bell. *Reviews of Modern Physics*, 65(3):803–815.

[61] Miller, J., Sanders, S., and Miyake, A. (2017). Quantum supremacy in constant-time measurement-based computation: A unified architecture for sampling and verification. *Physical Review A*, 96(6):062320.

[62] Morimae, T. (2017a). Hardness of classically sampling the one-clean-qubit model with constant total variation distance error. *Physical Review A*, 96(4):040302.

[63] Morimae, T. (2017b). Hardness of classically sampling the one-clean-qubit model with constant total variation distance error. *Physical Review A*, 96(4):040302.

[64] Morimae, T., Fujii, K., and Fitzsimons, J. F. (2014). Hardness of classically simulating the one-clean-qubit model. *Physical review letters*, 112(13):130502.

[65] Nebe, G., Rains, E. M., and Sloane, N. J. (2001). The Invariants of the Clifford Groups. *Designs, Codes, and Cryptography*, 24(1):99–122.

[66] Nielsen, M. and Chuang, I. (2010). *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniv edition.

[67] Nielsen, M. A. and Chuang, I. (2002). Quantum computation and quantum information.

[68] Novo, L., Bermejo-Vega, J., and García-Patrón, R. (2019). Quantum advantage from energy measurements of many-body quantum systems. Technical report.

[69] Papadimitriou, C. H. (1994). *Computational complexity*. Addison-Wesley.

[70] Pashayan, H., Bartlett, S. D., and Gross, D. (2017). From estimation of quantum probabilities to simulation of quantum circuits.

[71] Pashayan, H., Bartlett, S. D., and Gross, D. (2020). From estimation of quantum probabilities to simulation of quantum circuits. *Quantum*, 4:223.

[72] Pashayan, H., Wallman, J. J., and Bartlett, S. D. (2015). Estimating Outcome Probabilities of Quantum Circuits Using Quasiprobabilities. *Physical Review Letters*, 115(7).

[73] Peres, A. (1990). Incompatible results of quantum measurements. *Physics Letters A*, 151(3-4):107–108.

[74] Poulin, D., Blume-Kohout, R., Laflamme, R., and Ollivier, H. (2004). Exponential speedup with a single bit of quantum information: Measuring the average fidelity decay. *Physical review letters*, 92(17):177906.

[Preskill] Preskill, J. Lecture notes for Physics 219/Computer Science 219 Quantum Computation.

[76] Raussendorf, R., Browne, D. E., and Briegel, H. J. (2003). Measurement-based quantum computation on cluster states. *Phys. Rev. A*, 68:022312.

[77] Raussendorf, R., Browne, D. E., Delfosse, N., Okay, C., and Bermejo-Vega, J. (2017). Contextuality and Wigner function negativity in qubit quantum computation. *Physical Review A*, 95(052334).

[78] Rocchetto, A. (2018). Stabiliser states are efficiently PAC-learnable. *Quantum Information and Computation*, 18(7&8).

[79] Shor, P. W. and Jordan, S. P. (2007). Estimating jones polynomials is a complete problem for one clean qubit. *arXiv preprint arXiv:0707.2831*.

[80] Spekkens, R. W. (2008). Negativity and contextuality are equivalent notions of nonclassicality. *Physical Review Letters*, 101(2).

[81] Terhal, B. M. and DiVincenzo, D. P. (2002). Classical simulation of noninteracting-fermion quantum circuits. *Physical Review A - Atomic, Molecular, and Optical Physics*, 65(3):10.

[82] Terhal, B. M. and DiVincenzo, D. P. (2004). Adaptive Quantum Computation, Constant Depth Quantum Circuits and Arthur-Merlin Games. *Quantum Information & Computation*, pages 134–145.

[83] Valiant, L. G. (1984). A theory of the learnable. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 436–445. ACM.

[84] Van Den Nest, M. (2013). Universal quantum computation with little entanglement. *Physical Review Letters*, 110(6).

[85] Van den Nest, M. (2013). Universal quantum computation with little entanglement. *Physical review letters*, 110(6):060504.

[86] Veitch, V., Ferrie, C., Gross, D., and Emerson, J. (2012). Negative quasi-probability as a resource for quantum computation. *New Journal of Physics*, 14.

[87] Veitch, V., Wiebe, N., Ferrie, C., and Emerson, J. (2013). Efficient simulation scheme for a class of quantum optics experiments with non-negative Wigner representation. *New Journal of Physics*, 15.

[88] Vidal, G. (2003). Efficient classical simulation of slightly entangled quantum computations. *Physical Review Letters*, 91(14).

[89] Watts, A. B., Kothari, R., Schaeffer, L., and Tal, A. (2019). Exponential separation between shallow quantum circuits and unbounded fan-in shallow classical circuits. In *Proceedings of the Annual ACM Symposium on Theory of Computing*, pages 515–526. Association for Computing Machinery.

[90] Werner, R. F. (1989). Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. *Physical Review A*, 40(8):4277.

[91] Yoganathan, M. and Cade, C. (2019). The one clean qubit model without entanglement is classically simulable.

[92] Yoganathan, M., Jozsa, R., and Strelchuk, S. (2019). Quantum advantage of unitary Clifford circuits with magic state inputs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 475(2225).

[93] Yoran, N. and Short, A. J. (2006). Classical simulation of limited-width cluster-state quantum computation. *Physical Review Letters*, 96(17).